The phylogenetic position of *Cladiucha* within Tenthredinidae based on comprehensive mitochondrial phylogenomics and the evidence from comparative analyses of rRNA secondary structure

Gengyun Niu [1], Sijia Jiang [1], Özgül Doğan[2], Ertan Mahir Korkmaz[2], Mahir Budak[2], Duo Wu [1], Meicai Wei[1*]

1 College of Life Sciences, Jiangxi Normal University, Nanchang 330022, China; *

2 Faculty of Science, Department of Molecular Biology and Genetics, Sivas Cumhuriyet University, Sivas 58140, Turkey

Meicai Wei supervised the project, conceived the original idea. Gengyun Niu conceived and designed the experiments. They constantly optimized and adjusted the experimental plan.

Sijia Jiang performed heterogeneity analysis and constructed phylogenetic relationships. Duo Wu collected and analyzed the data. Özgül Doğan and Ertan Mahir Korkmaz designed the experimental plan to estimate the divergence time and carried out the experiment under the guidance of Mahir Budak. Sijia Jiang and Duo Wu prepared figures and/or tables, Gengyun Niu further drawn the final layout.

Gengyun Niu took the lead in writing the manuscript. Sijia Jiang wrote part of the manuscript with support from Özgül Doğan, and Ertan Mahir Korkmaz.

All authors provided critical feedback and helped shape the research, analysis, and manuscript.

*This work is supported by the National Social Science Foundation of China (Grant No. 31501885, 31672344).

*Corresponding author Meicai Wei (E-mail: weim@jxnu.edu.cn).

Abstract

Two mitogenomes of *Cladiucha* were newly reported and showed typical pattern of gene arrangement. The phylogenetic position of *Cladiucha* was obtained from tree reconstruction using various data treatment methods and substitution models. Significant heterogeneity in the nucleotide composition and mutational biases was found in the mitochondrial protein-coding genes, and the third codon position exhibited high levels of saturation. Therefore, 14 datasets were conducted under both site-homogeneous and site-heterogeneous models. The following conclusions were drawn from the phylogenetic analyses: (i) the monophyly of Tenthredinidae was confirmed, (ii) the monophyly of Allantinae + Tenthredininae + Megabelesinae was approved, and (iii) within the family, ((((Tenthredininae + Fenusinae) + Allantinae) + Megabelesinae) + Nematinae) is probably the most acceptable cladogram for the phylogeny of Tenthredinidae, which is also supported by the morphological analysis and a comparative study on the rRNA secondary structure. Divergence time estimation analyses indicated that diversification of the major superfamilies of the suborder Symphyta occurred around 232.9 Ma, and the splits of Tenthredinidae were dated to 146 Ma, which corresponds to the origin of the earliest lineages of flowering plants and major diversifications of core angiosperms, respectively. The *Cladiucha* arose in the Mid-Miocene; at that time, magnoliids are rapidly undergoing genus-species differentiation.

Key words: Mitochondrial phylogenomics; PhyloBayes; Secondary strucures; *Cladiucha*; Tenthredinidae

1. Introduction

Tenthredinidae is the largest and a considerably complex family of the paraphyletic suborder Symphyta (Hymenoptera) with 5645 extant species in 400 genera[1]. This family has been variously divided into five[2], six[3], seven[4,5], eight[6,7], 10[8] and 11[9] subfamilies, or into four families and 12 subfamilies [10] or six families and 17 subfamilies[11]. The numbers of the subfamilies within Tenthredinidae and their contents are not quite consistent, though some tribes have been recognized in most systematic research studies. *Megabeleses* Takeuchi, 1952[4] and *Cladiucha* Konow, 1902[12] are the only two genera among Tenthredinidae members of which the species are known to feeding on plants of Magnoliaceae, a basal lineage of Angiosperms. Benson[5] set a monotypic tribe, Cladiuchini, for *Cladiucha* Konow under Emphytinae (=Allantinae). Takeuchi[4] and Abe & Smith[7] placed the two genera into Allantinae of Tenthredinidae. Wei[13] pointed out that *Megabeleses*, *Cladiucha*, and two additional new genera represented a unique lineage among the Tenthredinidae and set a subfamily Megabelesesinae (= Megabelesinae) to place them, which might locate between Allantinae and Tenthredininae. This opinion was followed in the new system provided by Wei & Nie[11]. However, Taeger et al. [3] still placed the four genera into Allantinae. This systematic inconsistency indicates the requirement of more detailed studies to better understand their evolutionary history by using comprehensive molecular data under the current approaches.

The mitogenome data has recently been a widely applied tool in resolving longstanding questions about the evolutionary history of organisms due to their rapidly increasing number conjoint with their relative compactness in size and structure, inheritance type, absent or very infrequent recombination, and high mutation rate[14] [14]. But then, the presence of high A+T content, lineage-specific compositional heterogeneity, within-site rate variation (heterotachy) as well as nucleotide substitution saturation can lead to challenges in phylogeny reconstructions[15-18]. These undesirable features can frequently cause the forming of analytical artifacts such as long-branch attraction and deceptive phylogenetic relationships. The occurrence of analytical objects is a commonly encountered problem in the phylogeny reconstruction of intra- and interlineages of insect orders, most notably in Hymenoptera[19].

The effects of these biases in tree reconstruction are reduced using a variety of approaches, such as a series of data coding schemes and more sophisticated evolutionary model settings. Comparing the results of different data coding regimes is widely used to improve the signal-to-noise ratio in mitogenome data[17]. In general, misleading signals from protein-coding genes (PCGs) are not produced equally in the first, second, and third codon positions. In fact, the evolutionary rate of the third codon position is higher than that in the first and second codon positions. In some studies, the third codon positions have been removed from datasets. In addition, purine-pyrimidine (RY) coding (A&G→R; C&T→Y) in the third codon position is a common data coding to alleviate the effects of saturated synonymous nucleotide substitutions and rate heterogeneity among codon positions[15,16,20]. However, this position may provide useful phylogenetic signals in some groups. Removing or replacing sites may confound the phylogenetic relationship to some extent[16,17,20]. Regier et al.[21] and Zwick et al.[22] proposed and designed a new data treatment, Degen coding, to reduce nucleotide compositional heterogeneity and improve the resolution of deep-level arthropod relationships. This data scheme can retain nonsynonymous changes at the third codon position while eliminating all of the synonymous changes by extending other coding schemes to fully degenerate all of the codons[17,20].

In contrast to the wide application of PCGs in constructing phylogenetic relationships, rRNAs are used with caution. It has become a consensus, that rRNA could not be the protagonist in the inference, due to the ineffectiveness of aligning. Even using the alignments guidance by the secondary structure, few works could obtain satisfactory results. Due to insufficient data accumulation, homology structural motif have been explored in only a few taxa [23].

Here, we sequenced and reported two mitogenome of *Cladiucha*. We compiled 14 concatenate datasets and applied several commonly used phylogenetic inference methods based on both heterogeneity and homogeneity models to overcome the systematic bias arising from nonstationarity. We also estimated a time-frame of Symphyta evolution, especially the origin and diversification of Tenthredinidae. Finally, we discussed the application of the autapomorphy of rRNAs in the phylogeny frame. Our objective was to investigate the effect of compositional heterogeneity on phylogenetic reconstruction and usage of autapomorphy of the secondary structure of rRNAs within the rapid radiation of Tenthredinidae.

2. Methodology

DNA library construction and sequencing

Total DNA was extracted from *Cladiucha magnoliae*, and *C. punctata* using an E.Z.N.A.® Tissue DNA Kit (Omega, Norcross, GA) and was stored at -20°C according to the manufacturer's instructions. Sequencing libraries with approximately 250-bp insertions were constructed using a NEXT flex™ Rapid DNA-Seq Kit (Illumina, San Diego, CA) in accordance with the manufacturer's protocol. Each library was sequenced using an Illumina HiSeq 4000 to generate 150-bp paired-end reads at Shanghai Majorbio Bio-pharm Technology Co., Ltd. The genomic DNA yielded a total of 1.32G raw reads (SRR9998491) and 6.24G raw reads (SRR11177465).

Mitochondrial genome assembly and annotation

The *C. magnoliae* and *C. punctata* reads were imported into GENEIOUS R11 (http://www.geneious.com)[24] and assembled into contigs. The mitogenomes of *Megabeleses magnoliae* (unpublished), *M. liriodendrovorax* (unpublished), *Tenthredo tienmushana* (KR703581) and *Allantus luctifer* (KJ713152) were used as references with 'medium-low sensitivity' parameters. *M. magnoliae* was the primary reference among these mitogenomes when concerning *rrnS*. MITOS (http://mitos.bioinf.uni-leipzig.de/index.py) was used to predict the protein-coding, transfer RNA, and ribosomal RNA genes with annotation from a reference mitogenome. The total tRNA genes were identified by MITOS[25] using the invertebrate mitochondrial genetic code. Geneious v11.0.3 (http://www.geneious.com) was used to determine the initiation and termination codons of the PCGs referring to the sequences of other symphytan species with subsequent manual adjustment. Both the 3' and 5' ends of *rrnS* were annotated by secondary structure comparison, rather than by the flanking genes.

Saturation, nucleotide compositions and heterogeneity

MEGA v7.0 [26] was used to calculate the A+T content of the nucleotide sequences and the relative synonymous codon usage (RSCU). In the strand encoding the majority of the PCGs, the strand asymmetry was calculated using the following formulas: GC-skew = (G–C) / (G+C) and AT-skew = (A–T) / (A+T). Bubble charts (Fig. 1) were generated using JMP. Three-dimensional scatter plots of the A+T- and GC-skews and A+T% of four datasets (Fig. 7) were plotted using SigmaPlot 14®.

The disparity index (I_D) [27]detects differences in the evolutionary patterns between a pair of sequences, thereby indirectly measuring the level of base compositional heterogeneity. P values [28] smaller than 0.05 were considered significant. The I_D values were calculated for six nucleotides datasets (PCG123, PCG12, PCG1, PCG2, and PCG3) and an amino acid dataset (AA) using 1000 Monte Carlo replications, as implemented in MEGA 7.0 [26].

AliGROOVE[29] was used to analyze the heterogeneity of sequence divergence within various concatenated datasets for a default sliding window size. Indels in the nucleotide dataset were treated as ambiguities, and the BLOSUM62 matrix was used as the default amino acid substitution matrix.

The saturation level in the nucleotide and amino acid sequences was visualized and assessed using DAMBE [30] and ASaturA [31] respectively.

Secondary structure prediction

Secondary structures of ribosomal RNA (*rrnS* and *rrnL*) were predicted and inferred from alignments with models *Tenthredo tienmushana* (GenBank accession No. KR703581) and *Allantus luctifer* (GenBank accession No. KJ713152). First, the primary sequence and the secondary structure of the model species were aligned in MARNA[32] to output a consensus sequence and a consensus structure. This output file was then imported into SSU-ALIGN[33] to predict the rRNA secondary structures of *C. magnoliae* and *C. punctata*. Lastly, minor changes were made to transform the primary results into relative secondary structures manually. VARNA v3-93[34] and RnaViz 2.0.3[35] were used to draw the secondary structures of *rrnL* and *rrnS*. Helix numbering, with minor modifications, was performed, following the *Apis mellifera* rRNA secondary structures[36]. the RNA structural logo shown in Figures 3 and 4 were generated by RNALogo[37].

Data sets

The ingroup taxa for the phylogenetic analyses comprised 40 symphytan species, including eight taxa representatives of Tenthredinidea (Supplementary Table 1). Four taxa were also included as outgroups.

A total of 13 PCGs were aligned individually, excluding the stop codons by MAFFT v7[38], in PhyloSuite v1.1.5[39]. Amino acid alignment of the 13 PCGs was performed using Clustal X. The two ribosomal RNAs were aligned by MAFFT v7.

To eliminate the effect of saturation and compositional heterogeneity on the phylogenetic reconstruction, the following 14 concatenate datasets were compiled: 1) **PCG**: 13 protein-coding genes; 2) **PCG12**: 13 protein-coding genes excluding third codon positions; 3) **PCG12RY**: 13 protein-coding genes with RY coding strategy; 4) **PCGDegen**: 13 protein-coding genes with Degen coding strategy; 5) **9PCG**: nine protein-coding genes excluding four saturated genes; 6) **9PCG+4PCG12RY**: nine protein-coding genes and four saturated genes with RY coding strategy; and all six data sets above combined with two RNAs: 7) **PCG+RNA**; 8) **PCG12+RNA**; 9) **PCG12RY+RNA**; 10) **PCGDegen+RNA**; 11) **9PCG+RNA**; 12) **9PCG+4PCG12RY+RNA**; and two amino acid data sets: 13) **AA**: 13 protein-coding genes translated into amino acids; 14) **11AA**: unsaturated 11 amino acids.

Phylogenetic analysis

The phylogenies were estimated using the Maximum Likelihood (ML) and Bayesian Inference (BI) methods. The ML analyses were conducted using an IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/) and RAxML v8.2.10[40]. The Bayesian analyses were performed using MrBayes v3.2.6[41] and PhyloBayes v3.3[42].

Before performing the model-based phylogenetic analyses, we determined the best partitioning scheme and model using PartitionFinder v1.1.1[43,44]. The branch lengths and search algorithm settings were “unlinked” and “greedy”. The Bayesian information criterion (BIC) scheme and the corrected Akaike information criterion (AIC) were selected for the BI and ML methods, respectively. Accordingly, the datasets were partitioned into up to 16 subsets, and the best models were used for subsequent phylogenetic analyses.

The ML tree was calculated with branch support estimated from 500 bootstrap replicates. The BI analyses consisted of four separate runs of four chains each, for 10 million generations, sampling every 1000 trees and discarding the first 25% generations as burn-in. An additional Bayesian analysis was undertaken in PhyloBayes 3.3[42] under the heterogeneous model CAT-GTR for 9PCG and CAT-MtArt for 11AA, including four independent chains. These analyses were terminated after the likelihood of the sampled trees had stabilized, and the chains had satisfactorily converged (maxdiff < 0.3). The phylogenetic trees were visualised by FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Time estimation

Bayesian estimation of divergence times was performed using the dataset of 11AA in MCMCTree implemented in the PAML package v4.9[45] using the approximate likelihood method[46]. BASEML was used for the estimation of substitution rate per site, and this rate was used as the prior for

Bayesian analysis. MCMC was run by 50 x 10,000 iterations with the mtArt substitution model. The soft bound of the split of the superfamily Tenthredinoidea from others [between 260–290 million years ago (Ma)] was used as external secondary calibration based on previous studies[47,48]. The inferred nodal age of holometabolous insects (between 300-360 Ma) was used for the calibration of the root[49,50].

All of the related files have been uploaded to Figshare (https://figshare.com/account/home#/projects/78339)

3. Results

Cladiucha mitogenomes

The nearly complete mitogenomes of *C. magnoliae* and *C. punctate* are 15,214 bp and 15,301 bp in length, respectively, and are accessed in GenBank (accession number: MT295305-MT295306). The organizations of these two mitogenomes are summarized in Table 1.

Compared with the putative ancestral mitogenome of insects, only tRNA rearrangement events were detected. The position between *trnM* and *trnI* was swapped, and *trnR* was translocated from the second position in the *trnA-trnR-trnN-trnS1-trnE-trnF* cluster to the penultimate position. The A+T-rich region was partly sequenced (Fig. 5).

The gene overlaps, and gene gaps between *C. magnoliae* and *C. punctate* occurred in the same position. We identified six overlapping positions and 18 intergenic spacers. Twelve pairs of genes were directly adjacent. The total overlap length was 16 bp in both *C. magnoliae* and *C. punctata*, with each overlap ranging from 1 to 7 bp. The longest overlapping nucleotide was found between *atp6* and *atp8*. There were a total of 370 bp and 442 bp intergenic spacer sequences in *C. magnoliae* and *C. punctata*, respectively, which ranged from 1 to 151 bp in size. The longest intergenic spacer sequences for *C. magnoliae* and *C. punctata* of 130 bp and 151 bp, respectively, were located between *trnV* and *rrnS*. Using homologous searches, no significant similarities were found among the longest noncoding region in *C. magnoliae*, *C. punctata*, or other identified symphytan species.

Nine PCGs are located on the majority strand (J-strand), and four PCGs are located on the minority strand (N-strand) among the 13 PCGs of *Cladiucha* mitogenomes. The total lengths of the PCGs of these two species were 11,215 bp and 11,168 bp, respectively, accounting for 73.7% and 73.0%, respectively, of the entire genomes. All of the PCGs started with ATN codons, including four genes (*nad2*, *atp6*, *cox3*, and *cob*) starting with ATG, five genes (*cox1*, *cox2*, *atp8*, *nad5*, and *nad4L*) starting with ATT and four genes (*nad3*, *nad4*, *nad6*, and *nad1*) starting with ATA. The stop codons were mostly TAA, except for *nad4*, which had T as the stop codon.

The nearly complete mitogenomes of both *C. magnoliae* and *C. punctata* contain 22 tRNA genes. There are 14 and eight tRNAs encoded by the J and N strands, respectively. The position and orientation of the predicted tRNAs are identical in the two species, ranging in size from 64 bp (*trnD*) to 75 bp (*trnC*) (Table 1). Only *trnS1* (AGN) was missing a dihydrouridine (DHU) arm in both species. The predicted anticodons are identical to those of other known symphytan mitochondrial genomes (Fig. 6).

Both of the rRNA genes were encoded on the N-strand. The rRNAs *rrnL* were flanked by *trnL1* and *trnV*. *rrnS* were downstream of the *trnV*, with long intergenic space sequences. The length of the *rrnL* gene was 1344 bp in *C. magnoliae* and 1341 bp in *C. punctata*, with the same A+T contents of 84.7%. The lengths of the *rrnS* genes for *C. magnoliae* and *C. punctata* were 813 bp (84.0% A+T content) and 835 bp (84.4% A+T content), respectively. There are four domains with 27 helices in *rrnS* and six domains with 46 helices in *rrnL* in both species (Fig. 3-4).

Saturation, compositional heterogeneity and codon usage bias within nt data set

The nucleotide composition in the mitogenomes was described using two statistics: the A+T content and the AT- and GC-skews[51]. The AT-skew values of *C. magnoliae* and *C. punctata* were positive (0.0169 and 0.0145, respectively), whereas the corresponding GC-skew values were negative (-0.0183 and -0.1860) for the entire mitogenome. The total A+T content of *C. magnoliae* was 82.8%, ranging from 74.5% (*cox1*) to 90.1% (*atp8*), whereas that of *C. punctata* was 82.8%, ranging from 74.8% (*cox1*) to 91.4% (*atp8*). The highest A+T contents of 96.4% and 96.0% were found at the third codon position for *C. magnoliae* and *C. punctata*, respectively.

The RSCU of *C. magnoliae* and *C. punctata* are presented. A significant correlation has been found between codon usage and nucleotide composition in *C. magnoliae*, *C. punctata*, and also in other insect mitogenomes. UUU-Phe, UUA-Leu, AUU-Ile, AUA-Met, and AAU-Asn had the highest usage rate among all amino acids. The highest RSCUs were found for UUA-Leu (5.06 and 5.26 for *C. magnoliae* and *C. punctata*, respectively). CUC-Leu, CUG-Leu, GUC-Val, UCG-Ser, CCG-Thr, ACG-Thr, GCG-Ala, CGC-Arg, CGG-Arg, and AGC-Ser all have a high CG content and were least frequently used. Crozier and Crozier[52] discussed how the effect of amino acid occurrence depends on the nucleotide composition of codon usage, which can be calculated in terms of the ratio of the total occurrence of C+G (Pro, Ala, Arg, and Gly) to A+T (Phe, Ile, Met, Tyr, Asn, and Lys) in the mitogenomes. This ratio was 0.18 and 0.25 in *C. magnoliae* and *C. punctata*, respectively, which is close to that of other symphytan species (0.27-0.31)[53,54].

The total A+T% of PCGs + rRNA of all of the included hymenopteran species ranged between 83.40% (*Taeniogonalos taihorina*) and 75.13% (*Xyela curva*), with a mean of 79.16 (±2.18)%. Among Tenthredinidae, *Monocellicampa pruni* had the lowest A+T% (76.23%), whereas *Cladiucha* had the highest A+T% (82.25% and 82.15%) (Fig. 2A). Irrespective of whether the third codon positions (Fig. 2B) or the four oversaturated PCGs (Fig. 2C) were removed, *X. curva* and *T. taihorina* had the lowest and highest A+T contents, respectively, in Hymenoptera, and *M. pruni*, and the two *Cladiucha* species had the lowest and highest A+T contents, respectively, in Tenthredinidae. Counting only the 13 PCGs, the hymenopteran species with the lowest A+T content was *Orussus occidentalis* (74.12%), followed by *X. curva*, with 74.41% A+T (Fig. 2D); otherwise, the same data were obtained as for the three abovementioned datasets.

However, it was difficult to characterize each family based on the A+T content alone. Therefore, we drew three-dimensional scatter plots of the content, the AT-skew, and the GC-skew for 44 species in the PCG+RNA dataset. A radius was assigned to each coordinate point and used to simulate the locations of the relatively close species (Fig. 5). The different taxa occupied relatively independent positions in space, especially Xyelidae (Fig. 5B) and Pamphiloidae (Fig. 5C). In Tenthredinidae, *Cladiucha* occupied a relatively independent position in contrast to the other Tenthredinidae species, which can be confirmed in Fig. 7. As shown in the left of Fig. 7, compared to the rest Tenthredinidae species, *Cladiucha* had the highest A+T content overall and for each codon. When compared within Symphyta, *Cladiucha* had the highest AT-3%.

The I_D test showed different evolutionary patterns for all six datasets; that is, evolution was significantly nonhomogeneous.

A total of 946 pairwise comparisons were made for each dataset. The null hypothesis was rejected in 800 comparisons at the 5% significance level by comparing all 13 PCGs.

The number of rejected comparisons based on individual codon partitions suggested that the level of base compositional heterogeneity was the lowest in PCG2 (485), followed by PCG12 (677), PCG1 (681) and PCG3 (703). The maximum I_D was 19.59 in PCG1 (between *Taeniogonalos taihorina* and *Anopheles gambiae*) and 30.88 in PCG3 (between *C. magnoliae* and *Trachelus iudaicus*), suggesting higher evolutionary variability for the third codon than the first codon. The maximum I_D of 7.53 for the AA dataset was found between *T. taihorina* and *Anopheles gambiae*, which was below that for the 13 PCGs (27.076, between *Paroster microsturtensis* and *T. taihorina*) and confirmed that the amino acid characters were less compositionally heterogeneous than nucleotide characters. The most homogeneous data set corresponded to PCG2, which had the lowest maximum I_D (6.79, between *Parapolybia crocea* and *Neopanorpa phlchra*) and the fewest rejected comparisons (485). However, we were not able to use the PCG2-based phylogeny results to cover the Pamphiloidae.

The PCG3 mean ID was higher than that of PCG1 in each hymenopteran species, except for three Xyelidae species and two apocritan species. (Fig. 7). In addition to having the lowest and highest A+T contents, Tenthredinidae, *Cladiucha* and *Monocellicampa* had much higher nt123Sum and nt3Sum than the other species. Two *Xyela* species and *T. taihorina* in Apocrita had both nt12Sum and nt1Sum above 200 as well as the lowest and highest A+T contents in Symphyta, respectively.

The divergence between the nucleotide datasets (PCG123, PCG123RNA, PCG12, PCG12RNA, 9PCG123 and 9PCG123RNA) and the amino acid datasets (AA and 11AA) for the 44 species were analyzed using AliGROOVE software.

In the PCG dataset (Fig. 8A), the species with high heterogeneity were mainly distributed in Orussidae, Siricidae, and Cephidae, with one species each in Pergidae and Apocrita. However, the heterogeneity was reduced to varying degrees in the coding dataset (the PCG datasets with data recoding; Fig. 8 B-E). Relatively positive scores were obtained for the amino acid datasets for nearly all of the family comparisons. The 11AA datasets were the least likely to violate the phylogenetic assumptions. However, with the addition of RNA, except for the taxa above, almost all of the Cephidae showed heterogeneity (Fig. 8 F-H).

Comparative analysis of rRNA

In the *rrnL* secondary structure, H837 was a long variable stem with a small loop, whereas H991 and H1196 had variable helical lengths and loop sizes. The predicted structures of H563, H579, H777, H822, H1830, H2023, H2043, H2455, and H2547 were well conserved in *C. magnoliae*, *C. punctata*, and other reported insect species. H976 was redundant in two species and *T. tienmushana*. H1775 and H2347 were conserved with three pairs and a small loop. In the *rrnL* secondary structure of *C. magnoliae* and *C. punctata*, positions 293 and 294 (H777 helices) harbored AA, whereas *T. tienmushana* and *Birmella discoidalisa* (GenBank accession number KR703581, MF197548) harbored GU and *A. luctifer* (NC024664) harbored GA at the same site. The U at position 411 in H946 was replaced by A; the A-U (position 580 in H579) was changed to U-A; the G (position 1208 in H2588) was replaced by A; and the UU (position 1270-1271 in H2646) was changed to CC in *T. tienmushana*, *A. luctifer* and *B. discoidalisa*.

In the *rrnS* secondary structure, domain III was the most conserved domain in Tenthredinidae. H47 was a commonly observed variable loop, whereas H500, H769, H944, H1047, H1399 and H1506 were highly conserved in the two *Cladiucha* species and other symphytan species. In terms of sequence and structure, H9 and H367 were the most conserved helices in the two *Cladiucha* species and in *T. tienmushana*, *A. luctifer*, and *B. discoidalisa*. The H960, H577, and H1241 helices in the two *Cladiucha* species were more similar to *T. tienmushana* than *A. luctifer* and *B. discoidalisa*.

Phylogenetic relationships

The following conclusions are drawn from the results of 32 schemes (Fig. 9): (i) Xyelidae is the root of Hymenoptera, with Macroxyelinae and Xyelinae forming the monophylum. (ii) Within Tenthredinoidea, Argidae and Pergidae form a monophylum as sisters to the remaining Cimbicidae and Tenthredinidae. (iii) Within Tenthredinidae, Nematinae (*Analcellicampa xanthosoma* + *Monocellicampa pruni*) is a sister group of (Allantinae + Tenthredininae + Megabelesinae + Fenusinae) and the basal branch of Tenthredinidae.

However, the results of the 32 phylogenetic trees are inconclusive on the following issues: (i) the monophyly of Pamphilioidea; (ii) the relationships within Unicalcarida; and (iii) although most branches are highly supported, seven topologies have emerged in Tenthredinidae.

Eight schemes supported topology No. 3., which was considered to be optimal in this work and was supported by phylogenetic tree construction with only eight Tenthredinidae species. Within Tenthredinidae, all 13 PCGs were unsaturated. Therefore, eight taxa_PCG and eight taxa_PCG + rRNA datasets were used to build the BI tree, both of which supported topology No. 3. Moreover, this topology was confirmed by the rRNA secondary structure. Topology No. 1 confirmed to topology No. 3, except for the relative position of Megabelesinae and Allantinae.

Three schemes supported topology No. 5. and three other schemes, including the BI and ML trees of 11AA, supported topology No. 6. They both cover the monophyly of Allantinae + Tenthredininae + Megabelesinae. However, the secondary structure (Fig. 10) shows that Allantinae, Tenthredininae, and Fenusinae have multiple common derivatives, excluding topologies No. 5 and No. 6 to some extent.

Although there are nine schemes supporting topology No. 4 and six schemes supporting topology No. 2, respectively, no apomorphic morphological character was found to confirm this result.

Dated phylogeny

A chronogram of divergence times for the included hymenopteran species based on the obtained tree topology is shown in Fig. 11. According to the divergence time analysis, the crown age of Hymenoptera was estimated as 318.1 Ma [95% CI (300.88 – 349.61 Ma)] corresponding to Pennsylvanian (Carboniferous). Diversification of the major superfamilies of Symphyta occurred between the beginning of the Late Triassic (Carnian, 232.9 Ma) and Jurassic-Cretaceous transition (145.0 Ma). The splits of the families of the superfamily Tenthredinoidea were estimated as corresponding to Late Jurassic [Pergidae + Argidae, 159.5 Ma, 95% CI (94.73 – 224.25 Ma)] and Early Jurassic [Cimbicidae + Tenthredinidae, 187.5 Ma, 95% CI (137.98 – 238.00 Ma)]. The splits of the subfamilies Nematinae and Allantinae from other subfamilies of the family Tenthredinidae were dated to 146 Ma [95% CI (102.11 – 195.15 Ma)] and 116.3 Ma [95% CI (77.53 – 162,29 Ma)], coinciding to Late Jurassic (Tithonian) and Early Cretaceous (Aptian), respectively. The split time of the subfamily Tenthredininae from Heterarthrinae + Megabelesesinae was inferred as 103.6 Ma [95% CI (64.01 – 148.06 Ma)], corresponding to Early Cretaceous (Aptian). The divergence between the subfamilies Heterarthrinae and Megabelesesinae was estimated to take place in Late Cretaceous [Coniacian, 86.4 Ma]. The divergence time of *C. magnoliae* and *C. punctata* corresponds to the Mid-Miocene [15.5 Ma, 95% (5.7 – 34.39 Ma)].

4. Discussions

Mitogenome characteristics of *Cladiucha*

The translocation of *trnR* is a shared derived character of *Cladiucha*, which is a novel event in Symphyta and could be identified as a synapomorphy of *Cladiucha*. The *trnI*, *trnQ*, and *trnM* gene cluster rearrangement events are too random to reconstruct the pattern of genome rearrangements. Thus, the evidence from gene rearrangements could not be used to determine the phylogeny of Tenthredinidae.

In *C. magnoliae*, *C. punctata*, a conserved overlap and two conserved noncoding sequences were found. The conserved overlap (A) located between *atp8* and *cox3*, and the conserved noncoding sequences (AA and T) located between *nad4L* and *trnT*, and between *trnP*, and *nad6*. These characteristics were also found in *A. luctifer*, *T. tienmushana* and *B. discoidalisa*.

We found 21 and 20 mismatched pairs in the tRNAs of *C. magnoliae* and *C. punctata*, respectively, that were composed of G-U (10 pairs in *C. magnoliae* and 11 pairs in *C. punctata*), U-U (9 pairs in *C. magnoliae* and seven pairs in *C. punctata*) and G-A (two pairs in each species) mispairing, which is common in Hymenoptera[14,55-59]. The mismatched pairs were mainly located on the DHU and the anticodon stems.

The I_D corresponds to the disparity between the observed compositional difference for two sequences and the expected difference under homogeneity. That is, the I_D is zero when the homogeneity assumption is satisfied[27].

Homogeneity (I_D=0) was observed between *Cladiucha* and *Tremex*, *Trichiosoma*, and *Megalodontes* (the first and second datasets). However, these taxa seldom share biological or evolutionary characteristics. *Cladiucha* distributes in southern China and neighboring countries of southeastern Asia, whereas the other taxa distribute in northern China. The *Cladiucha* species are monophagous leaf feeders of *Magnolia* and *Manglietia* of Magnoliaceae, a primitive family of Angiospermae. *Tremex* larvae bore into the stems of many species of Salicaceae of Rosopsida and Betulaceae and Fagaceae of Hamamelidopsida. The *Trichiosoma* species feed on the leaves of *Salix* spp. and *Populus* spp. of Salicaceae, *Betula* spp. of Betulaceae and *Prunus* spp. and *Sorbus* spp. of Rosaceae of Rosopsida. The *Megalodontes* larvae feed on the leaves of *Sibiraea* spp. of Rosaceae. *Cladiucha* larvae are gregarious, whereas the other aforementioned larvae are all solitary.

Phylogenetic relationship

Different substitution patterns were observed among the Symphyta lineages, and all the datasets appeared to be biased. However, the AA dataset was the least likely to violate the phylogenetic assumptions, and the heterogeneity model performed best in restoring the monophyly of Apocrita and Hymenoptera and supporting the formation of the monophyletic group of Apocrita and the Orussidae. Thus, we used tree No. 3 as a reference phylogeny (Fig. 9)

Previous research studies have not considered the heterogeneity of the Symphyta sequence. Some of the novel methods that explicitly account for bias can overcome the yield problem caused by heterogeneity. However, discrepancies between such inferences within Tenthredinidae still exist.

The relationships between more heterogeneous taxa, such as Vespina, Siricoidea, Xiphydriidae, and Cephidae, have not been resolved in the homogeneity model. However, the heterogeneity model recovered Orussidae within Vespina with Apocrita.

All four schemes of the heterogeneity models support the monophyly of Pamphilioidea. However, the relationships between Cephidae remain unresolved.

Given that all three schemes of the heterogeneity model produced topology No. 2 of Tenthredinidae, we speculated that no advantage was obtained in using the heterogeneity model to determine the relationship between non-heterogeneous species.

The ML of all of the datasets did not support the monophyly of Pamphilioidea or perform better than BI in restoring the monophyly of Unicalcarida. In determining the internal relationship of Cephidae, the problems solved by BI for the same dataset could not be resolved by ML. Thus, we speculate that ML is inferior to BI in terms of tree construction, which is consistent with the results of many previous studies[60].

In contrast to the conclusions drawn from a study of Coleoptera[60], we found that the mitogenomic data containing RNA reduced the support of the Unicalcarida branches to varying degrees. The multiple branches inside the Cephidae also collapsed under RNA addition. We speculated that simply using the RNA primary sequence alignment results introduces more noise into the tree construction. Exploring the use of the secondary structure to construct phylogeny may lead to breakthroughs in the rational use of RNA genes.

Systematic placement of *Cladiucha*

Based on morphological phylogeny, Wei & Nie[11] proposed the Tenthredinoidea s. str. System, which should form a framework for subsequent research of Symphytan phylogeny. Some of the results were subsequently confirmed by molecular data[61], morphological data[62,63], or combinations [64]. Nevertheless, *Cladiucha* has not sampled in the cladistic framework except for Wei & Nie[11].

Wei & Nie (1998) categorized *Cladiucha* as neither Allantinae nor Tenthredininae and established Megabelesinae, which included *Cladiucha*, *Megabeleses*, and two relative genera, based on morphological phylogeny. Our mitochondrial genomic data place *Cladiucha* as a sister to either Tenthredininae (topology No. 5) or Fenusinae (topology No. 2 and No. 4).

The monophyly of Allantinae + Tenthredininae + Megabelesinae (topology No. 5 and No. 6) is strongly supported by many apomorphic characters, such as a derived upper head (for example, the differentiated frons), a large and broad mesepimeron covering the post-thoracic spiracle, a long and multisegmented lancet and a lance with a very short radix. However, the evidence from rRNA went against it. The violation may have occured due to the limited taxon sampling.

Placing *Cladiucha* as a sister to Allantinae is not supported by our results. In addition to the mitochondrial genome data, the following evidence supports this result. The adult morphological characters support that Allantinae is a sister group of Tenthredininae; for example, both groups share derived mandibles, an enlarged head and a broadened anterior lobe of the pronotum, whereas, in Megabelesinae, the mandibles are simple and bidentate, the head is not distinctly enlarged behind the eyes, and the anterior lobe of the pronotum is very narrow. Megabelesinae species share a peculiar apomorphic character: the female lance has a long and broad membranous lobe and approaches the upper 0.3–0.5 of the lancet, which supports that Megabelesinae may have been differentiated and isolated from the Allantinae-pattern ancestor over a considerably long period.

Dated phylogeny of Symphyta

Li et al.[65] depicted the phylogeny of angiosperm, with nearly 3000 chloroplast genomes from species representing all 64 orders. Molecular clock dating supports the earliest origin of flowering plants, dating the crown group to the Later Triassic period (~209 Ma), which corresponds with the timing of the diversification of the largest superfamily in Symphyta, Tenthredinoidea (marked as a green block in Fig. 11). Subsequently, the occurrence of the spectacular diversification of core angiosperms may arise during the early stages of mesangiosperm evolution (~164 Ma-159 Ma), which coincides with the rapid radiation of Tenthredinidae, which is the most diverse lineage in Symphyta.

All of the known host plants of the Megabelesinae species are plants of Magnoliaceae, a very primitive angiosperm family. The magnoliid orders diverged in the Lower Cretaceous (~140-132 Ma), as estimated in Li et al.[65], or even older[66]. Either the former or the latter is far earlier than the split of Megabelesinae with others. This once again proved that more taxon needs to be sampled in the phylogeny. The species differentiation within *Cladiucha* corresponds with the genus and species differentiation within Magnolia[67] (marked as a green block in Fig. 11).

5. Conclusions

1. tRNA gene rearrangements have been found in the mitochondrial genomes of Cladiucha. Two species both have the highest A+T content in the Tenthredinidae and, counting only the third codon, in the Symphyta. However, neither the ID nor the AliGROOVE showed heterogeneity.

2. Heterogeneity models can better describe the kinship between heterogeneous species. However, there is no obvious substantive improvement in analyzing the relationship of the species that do not show the heterogeneity.

3. The conserved motif in rRNAs could serve as a better potential synapomorphy than the optimized alignment.

4. Tenthredinidae is a monophyletic group. Within this family, ((((Tenthredininae + Fenusinae) + Allantinae) + Megabelesinae) + Nematinae) is probably the most acceptable cladogram for the

phylogeny of Tenthredinidae, which is also supported by a comparative analysis of the RNA secondary structure.

ACKNOWLEDGEMENTS

The members of the Lab of Insect Systematics and Evolutionary Biology (LISEB) from Jiangxi Normal University are thanked for their contributions to laboratory work. We thank all the reviewers for their comments.

References

[1] Taeger A E A. ECatSym – Electronic World Catalog of Symphyta (Insecta, Hymenoptera). [M]. Senckenberg Deutsches Entomologisches Institut (SDEI), 2018.
[2] Malaise R. Tenthredinoidea of South-Eastern Asia with a general zoogeographical review [J]. Opuscula Entomologica, 1945, 4: 1–288.
[3] Taeger A, Blank S M, Liston A D. World catalog of Symphyta (Hymenoptera)[J]. Zootaxa, 2010, 2580: 1–1064.
[4] Takeuchi K. A generic classification of the Japanese Tenthredinidae (Hymenoptera: Symphyta)[M]. Kyoto: 1952: 1–90.
[5] Benson R B. On the Classification of Sawflies (Hymenoptera Symphyta)[J]. Transactions of the Royal Entomological Society of London, 1938, 87(15): 353–384.
[6] Ross H H. Suborder Symphyta (= Chalastogastra) [except the Siricoidea, the Pamphiliidae, and the genus Periclista][M]. Washington: United States Department of Agriculture Agriculture Monograph, 1951.
[7] Abe M, Smith D R. The genus-group names of Symphyta (Hymenoptera) and their type species[J]. Esakia, Fukuoka, 1991, 31: 1–115.
[8] Ross H H. A Generic Classification of the Nearctic Sawflies (Hymenoptera, Symphyta)[M]. 15. Urbana: Illinois biological monographs, 1937: 1–173.
[9] Rohwer S A. Technical papers on miscellaneous forest insects. II. The genotypes of the sawflies or woodwasps, or the superfamily Tenthredinoidea.[M]. 20. Washington, DC: Technical series / US Department of Agriculture, Bureau of Entomology, 1911. 60-109.
[10] Ashmead W H. Classification of the Hornbills and Sawflies, or the Suborder Phytophaga[J]. The Canadian Entomologists, 1898, 30(6): 141–145, 177–183, 225–232, 249–257, 281–287, 305–316.
[11] Wei M, Nie H. Generic list of Tenthredinoidea s. str. in new systematic arrangement with synonyms and distribution data[M]. 18. Zhuzhou: Journal of Central South Forestry University, 1998. 23-31.
[12] Konow F. Neue Blattwespen. (Hym.). [M]. 2. Teschendorf bei Stargard i. Mecklenburg, 1902.
[13] Wei M. A new subfamily and two new genera of Tenthredinidae (Hymenoptera: Tenthredinomorpha)[J]. Entomotaxonomia, 1997, 19(suppl.): 69–76.
[14] Cameron S L. Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny[J]. Annual Review of Entomology, 2014, 59(1): 95–117.
[15] Yang H, Li T, Dang K, et al. Compositional and mutational rate heterogeneity in mitochondrial genomes and its effect on the phylogenetic inferences of Cimicomorpha (Hemiptera: Heteroptera)[J]. BMC Genomics, 2018, 19(1): 1–13.
[16] Timmermans M J T N, Barton C, Haran J, et al. Family-level sampling of mitochondrial genomes in Coleoptera: Compositional heterogeneity and phylogenetics[J]. Genome Biology and Evolution, 2016, 8(1): 161–175.
[17] Song F, Li H, Jiang P, et al. Capturing the phylogeny of holometabola with mitochondrial genome data and Bayesian site-heterogeneous mixture models[J]. Genome Biology and Evolution, 2016, 8(5): 1411–1426.
[18] Li H, Shao R, Song N, et al. Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences[J]. Scientific Reports, 2014, 5: 1–10.
[19] Talavera G, Vila R. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny[J]. BMC Evolutionary Biology, 2011, 11: 315.
[20] Hojun S, Nathan C, Sheffield, Stephen L, Cameron, et al. When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics[J]. Systematic Entomology 2010, 35: 429–448.
[21] Regier J C, Shultz J W, Zwick A, et al. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences[J]. Nature, 2010, 463(7284): 1079–1083.
[22] Zwick A, Regier J C, Zwick D, C. Resolving discrepancy between nucleotides and amino acids in deep-level arthropod phylogenomics: differentiating serine codons in 21-amino-acid models[J]. Plos one, 2012, 7(11): e47450.
[23] Song N, Lin A, Zhao X. Insight into higher-level phylogeny of Neuropterida: Evidence from secondary structures of mitochondrial rRNA genes and mitogenomic data[J]. PLoS One, 2018, 13(1): e0191826.
[24] Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data[J]. Bioinformatics, 2012, 28(12): 1647–1649.
[25] Bernt M, Donath A, Jühling F, et al. MITOS: improved de novo metazoan mitochondrial genome annotation[J]. Molecular phylogenetics evolution, 2013, 69(2): 313–319.
[26] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. Molecular biology evolution, 2016, 33(7): 1870–1874.
[27] Kumar S, Gadagkar S R. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences[J]. Genetics, 2001, 158(3): 1321–1327.
[28] North B V, Curtis D, Sham P C. A note on the calculation of empirical P values from Monte Carlo procedures[J]. The American Journal of Human Genetics, 2002, 71(2): 439–441.

[29] Kück P, Meid S A, Groß C, et al. AliGROOVE – visualization of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support[J]. BMC Bioinformatics, 2014, 15: 294.
[30] Xia X. DAMBE6: new tools for microbial genomics, phylogenetics, and molecular evolution[J]. Journal of Heredity, 2017, 108(4): 431–437.
[31] Van De Peer Y, Frickey T, Taylor J S, et al. Dealing with saturation at the amino acid level: A case study involving anciently duplicated zebrafish genes[J]. Gene, 2002, 295(2): 205–11.
[32] Siebert S, Backofen R. MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons[J]. Bioinformatics, 2005, 21(16): 3352-9.
[33] Nawrocki E P: Structural RNA homology search and alignment using covariance models[M]. Washington: University in St. Louis, 2009.
[34] Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure[J]. Bioinformatics, 2009, 25(15): 1974–1975.
[35] De Rijk P, Wuyts J, De Wachter R. RnaViz 2: an improved representation of RNA secondary structure[J]. Bioinformatics, 2003, 19(2): 299–300.
[36] Gillespie J J, Johnston J, Cannone J J, et al. Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of Apis mellifera (Insecta: Hymenoptera): structure, organization, and retrotransposable elements [J]. Insect Molecular Biology, 2006, 15(5): 657–686.
[37] Chang T H, Horng J T, Huang H D. RNALogo: a new approach to display structural RNA alignment[J]. Nucleic Acids Res, 2008, 36(Web Server issue): W91-6.
[38] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. Molecular biology and evolution, 2013, 30(4): 772–780.
[39] Zhang D, Gao F, Jakovlić I, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies[J]. Molecular ecology resources, 2020, 20(1): 348–355.
[40] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies[J]. Bioinformatics, 2014, 30(9): 1312–1313.
[41] Ronquist F, Teslenko M, Van Der Mark P, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space[J]. Systematic biology, 2012, 61(3): 539–542.
[42] Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating[J]. Bioinformatics, 2009, 25(17): 2286–2288.
[43] Lanfear R, Calcott B, Ho S Y, et al. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses[J]. Molecular biology and evolution, 2012, 29(6): 1695–1701.
[44] Guindon S, Dufayard J F, Lefort V, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0[J]. Systematic biology, 2010, 59(3): 307–321.
[45] Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood[J]. Computer Applications in the Biosciences Cabios, 1997, 13(5): 555-556.
[46] Dos Reis M, Yang Z. Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times[J]. Molecular Biology and Evolution, 2011, 28(7): 2161–2172.
[47] Peters R S. Evolutionary history of the Hymenoptera[J]. Current Biology, 2017, 27: 1–6.
[48] O'reilly J E, Dos Reis M, Donoghue P C J. Dating tips for divergence-time estimation[J]. Trends in Genetics, 2015: 637–650.
[49] Wiegmann B M. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects[J]. BMC Biology, 2009, 7(1): 34.
[50] Gaunt M W, Miles M A. An Insect Molecular Clock Dates the Origin of the Insects and Accords with Palaeontological and Biogeographic Landmarks[J]. Molecular Biology and Evolution, 2002, 19(5): 748–761.
[51] Perna N T, Kocher T D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes[J]. Journal of molecular evolution, 1995, 41(3): 353–358.
[52] Crozier R, H, Crozier Y, C. ThemitochondrialgenomeofthehoneybeeApis mellifera: complete sequence and genome organization[J]. Genetics, 1993, 133: 97–117.
[53] Korkmaz E M, Doğan Z, Budak M, et al. Two nearly complete mitogenomes of wheat stem borers, Cephus pygmeus (L.) and Cephus sareptanus Dovnar-Zapolskij (Hymenoptera: Cephidae): An unusual elongation of rrnS gene[J]. Gene, 2015, 558(2): 254–264.
[54] Du S, Niu G, Nyman T, et al. Characterization of the mitochondrial genome of Arge bella Wei & Du sp. nov. (Hymenoptera: Argidae)[J]. Peerj, 2018, 6: e6131.
[55] Wei S J, Niu F F, Du B Z. Rearrangement of trnQ-trnM in the mitochondrial genome of Allantus luctifer (Smith) (Hymenoptera: Tenthredinidae)[J]. Mitochondrial DNA Part A 2014: 856–858.
[56] Song S-N, Tang P, Wei S-J, et al. Comparative and phylogenetic analysis of the mitochondrial genomes in basal hymenopterans[J]. Scientific Reports, 2016, 6: 20972.
[57] Dowton M, Cameron S L, Dowavic J I, et al. Characterization of 67 mitochondrial tRNA gene rearrangements in the hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral[J]. Molecular Biology and Evolution, 2009, 26(7): 1607–1617.

[58] Doğan Ö, Korkmaz E M. Nearly complete mitogenome of hairy sawfly, Corynis lateralis (Brulle, 1832) (Hymenoptera: Cimbicidae): rearrangements in the IQM and ARNS1EF gene clusters[J]. Genetica, 2017, 145(4-5): 341–350.
[59] Castro L R, Dowton M. The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of Perga condei (Hymenoptera: Symphyta: Pergidae)[J]. Molecular Phylogenetics and Evolution, 2005, 34(3): 469–479.
[60] Yuan M L, Zhang Q L, Zhang L, et al. High-level phylogeny of the Coleoptera inferred with mitochondrial genome sequences[J]. Molecular Phylogenetics and Evolution, 2016, 104: 99–111.
[61] Malm T, Nyman T. Phylogeny of the symphytan grade of Hymenoptera: new pieces into the old jigsaw(fly) puzzle[J]. Cladistics-the International Journal of the Willi Hennig Society, 2014, 31(1): 1-17.
[62] Schulmeister S. Review of morphological evidence on the phylogeny of basal Hymenoptera, with a discussion of the ordering of characters[J]. Biological Journal of the Linnean Society, 2003, 79: 209-243.
[63] Lars V. Phylogeny and classification of the extant basal lineages of the Hymenoptera (Insecta)[J]. Zoological Journal of the Linnean Society, 2001, 131: 393-442.
[64] Schulmeister S. Simultaneous analysis of basal Hymenoptera (Insecta): Introducing robust-choice sensitivity analysis[J]. Biological Journal of the Linnean Society, 2003, 79(2): 245-275.
[65] Li H T, Yi T S, Gao L M, et al. Origin of angiosperms and the puzzle of the Jurassic gap[J]. Nat Plants, 2019, 5(5): 461–470.
[66] Beaulieu J M, O'meara B C, Crane P, et al. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms[J]. Systematic biology, 2015, 64(5): 869–878.
[67] Nie Z L, Wen J, Azuma H, et al. Phylogenetic and biogeographic complexity of Magnoliaceae in the Northern Hemisphere inferred from three nuclear data sets[J]. Molecular Phylogenetics and Evolution, 2008, 48: 1027–1040.

Figure Legends

Fig. 1 Bubble chart of the AT- and GC-skew and A+T% of mitochondrial genomes of Symphyta based on PCG+rRNA data set. (a) contains the bubble charts of 44 species in this study, and can be split into (b) to (f) according to different lineages; (b) contains Xyelidae (red) and Unicalcarida except for Cephidae (pink); (c) contains Pamphilioidea (orange) and the outer groups (purple); (d) contains Tenthredinidae (blue); (e) contains Argidae (light green), Pergidae (light green), Cimbicidae (deep green); (f) contains Cephidae (yellow). The letter above the dot represents the abbreviation of the species name.

Fig. 2 Three-dimensional scatter-plot of the AT- and GC-skew and A+T % of the mitochondrial genomes of Symphyta based on PCG+rRNA (A), PCG12 (B), 9PCG (C) and PCG (D) data sets. PCG + rRNA data set contains 13 protein coding genes and two rRNAs; PCG12 data set contains 13 protein coding genes (excluding the third site); 9PCG data set contains 9 protein coding genes (excluding 4 supersaturated protein coding genes); PCG data set contains 13 protein coding genes.

Fig. 3 Predicted rrnS secondary structure in the *Cladiucha* mitochondrial genome. The numbering of helices follows Gillespie et al. (2006). Roman numerals refer to domain names. Tertiary inter-actions and base triples are connected by continuous lines. *C. magnoliae* as a basemap and base change among *Cladiucha* species are presented in circles with red (*C. magnoliae*) and pink (*C. punctate*) color. Logo is generated from 8 species of Tenthredinidae (*C.punctate*, *C. magnoliae*, *Tenthredo tienmushana, Allantus luctifer, Asiemphytus rufocephalus, Analcellicampa xanthosoma, Birmella discoidalisa, Monocellicampa pruni*), each Logo graph is composed of stacks of letters, with one stack for each position in the consensus RNA secondary structure and the size of the letters represents the degree of similarity between bases.

Fig. 4 Predicted rrnL secondary structure in the *Cladiucha* mitochondrial genome. The numbering of helices follows Gillespie et al. (2006). Roman numerals refer to domain names. Tertiary inter-actions and base triples are connected by continuous lines. *C. magnoliae* as a basemap and base change among *Cladiucha* species are presented in circles with red (*C. magnoliae*) and pink (*C. punctate*) color. Logo is generated from 8 species of Tenthredinidae (*C. punctate, C. magnoliae*, *Tenthredo tienmushana, Allantus luctifer, Asiemphytus rufocephalus, Analcellicampa xanthosoma, Birmella discoidalisa, Monocellicampa pruni*), each Logo graph is composed of stacks of letters, with one stack for each position in the consensus RNA secondary structure and the size of the letters represents the degree of similarity between bases.

Fig. 5 Mitochondrial genome organization of *Cladiucha* with reference to the ancestral type of insect mitochondrial genomes. Genes transcribed from the J- and N-strands are shown in green and orange, respectively. The A+ T-rich region is indicated by blue, and tRNA genes are labeled by their single-letter amino acid code.

Fig. 6 Predicted secondary structures of 22 tRNA genes of *Cladiucha*. Dashes indicate Watson-Crick base pairs, and dots indicate G-U base pairing.

Fig. 7 A+T content and disparity index chart. The chart on the left shows the AT content of 44 species: AT% , AT-1%, AT-2% and AT-3%. On the other side of the chart is the AT content of the nucleotides and amino acid (nt123Sum, nt12Sum, nt1Sum, nt2Sum, nt3Sum and AA) for 44 species.

Fig. 8 Heterogeneity analysis of PCGs and PCGRNA datasets. AliGROOVE heat maps of pairwise sequence comparisons for the protein coding genes with the nucleotide datasets (PCG123, PCG123RNA, PCG12, PCG12RNA,9PCG123 and 9PCG123RNA) and the amino acid datasets (AA and 11AA) for the 44 species. The AIiGROOVE graph shows the mean similarity scores between sequences. AliGROOVE scores range from - 1 (indicating great difference in rates from the remainder of the data set, ie, red coloring implies the significant heterogeneity) to +1 (indicating rates match all other comparisons, ie, blue labeling).

Fig. 9 Phylogenetic hypothesis resulting under MtArt + CAT from Phylobayes analysis of 11AA data set. Branch colors summarize support values (see key). Box charts presented on backbone branches represent clade support in each scheme (see “Scheme support”). Methods and datasets of phylogenetic tree are summarized in the upper right corner of the figure. The dotted box refers to the 7 branching models reconstructed with different methods and datasets.

Fig. 10 Secondary structure drawings of the homology. The conserved structure of *rrnS* and *rrnL* is on the right side of each branch. The numbers in parentheses represent the number of homologues in the family. The homologous stem-loop structures on the branch are framed in the structural diagram.

Fig. 11 Dated phylogeny of Symphyta. The axis on the bottom refers to million years and shows the geological time. The blue bars on the nodes represent 95% of high posterior density of divergence times obtained from MCMCTree. The divergence times of each node obtained from MCMCTree analysis was written in black on the nodes. The red dots and the times written in red on the nodes indicate the divergence time obtained from MEGAX. The green and yellow shaded areas delineate the arising of Tenthredinoidea and *Cladiucha*, respectively, depending on the minimum or maximum ages. The phylogenetic trees given at the bottom are the angiosperm tree from Li et al (2019) and Magnolia tree from Nie et al., (2008), respectively.

Figures

Fig. 1

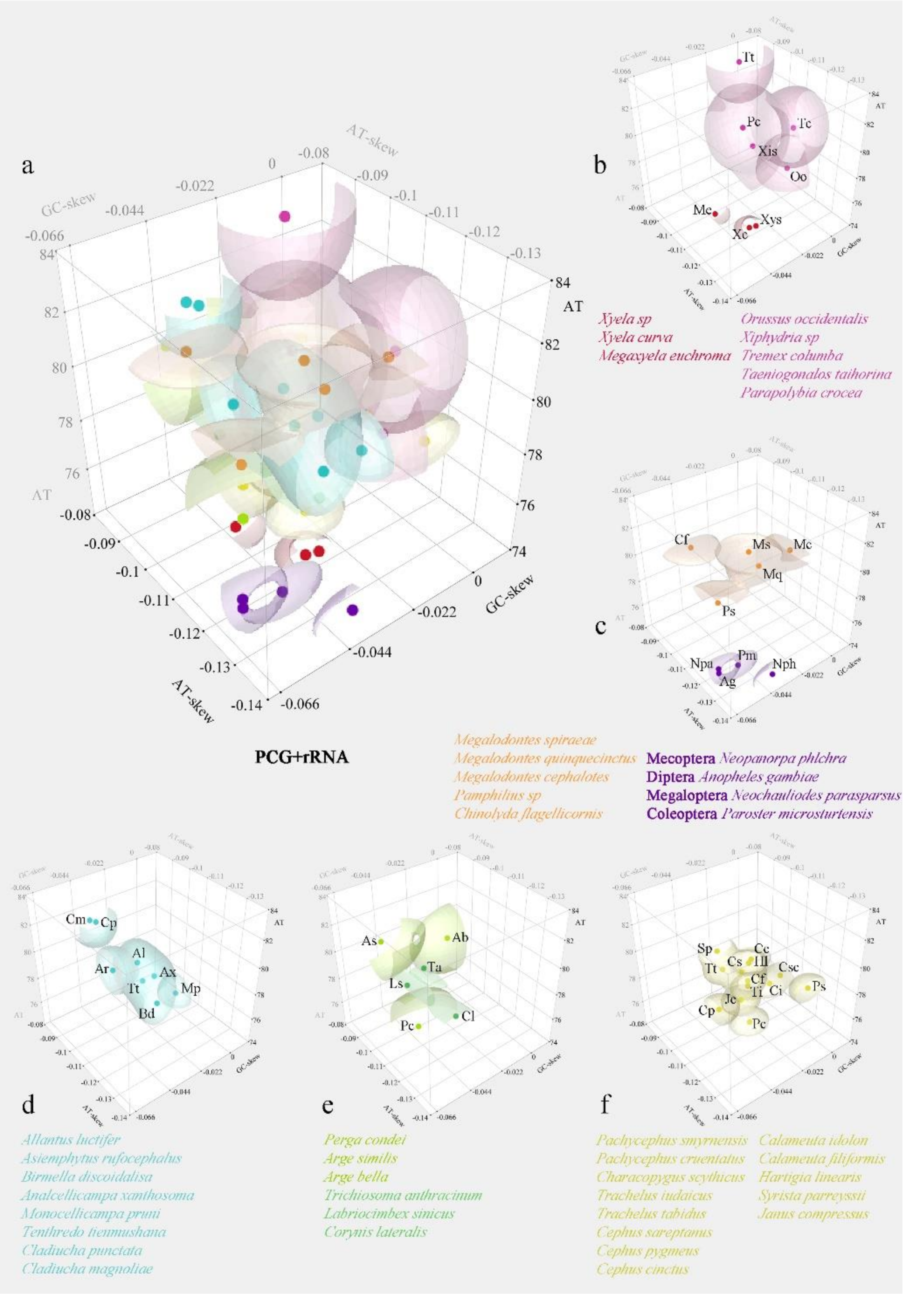


Fig. 2

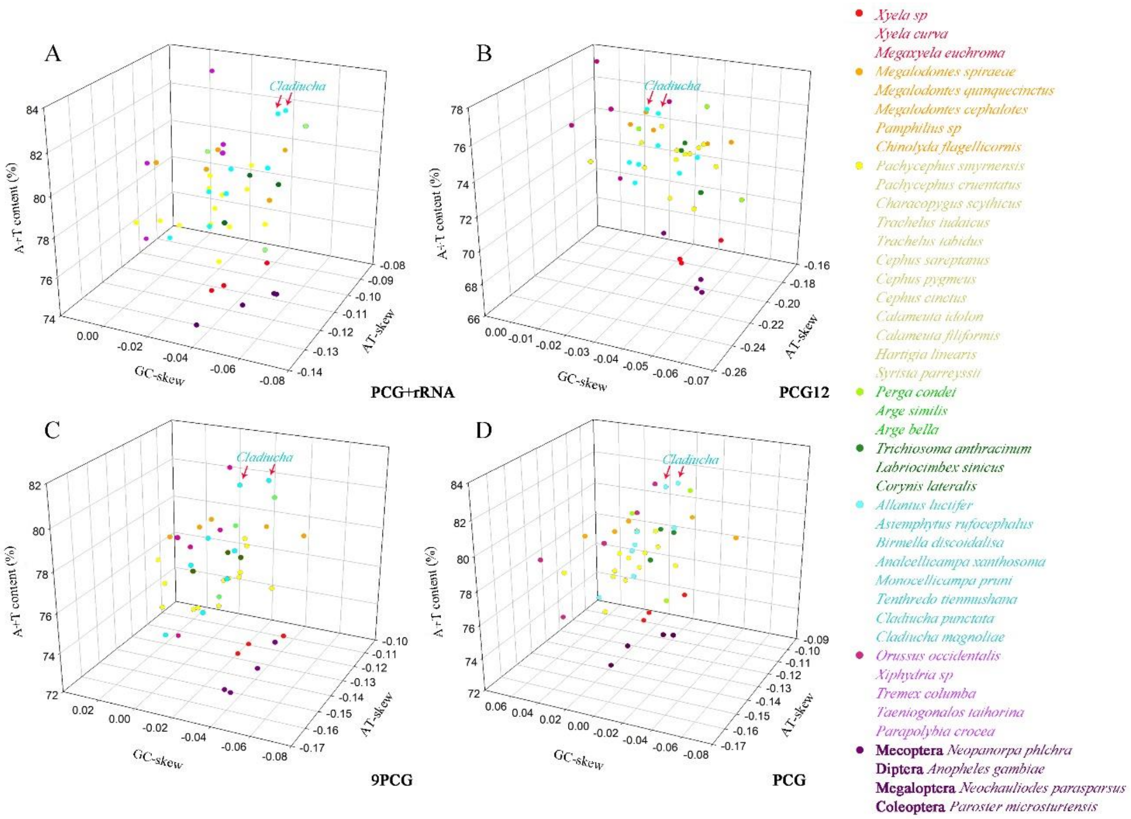

A
B
C
D
Claducha
A+T content (%)
GC-skew
AT-skew
84
82
80
78
76
74
0.00
-0.02
-0.04
-0.06
-0.08
-0.09
-0.10
-0.11
-0.12
-0.13
-0.14
PCG+rRNA
72
70
68
66
-0.01
-0.03
-0.05
-0.07
-0.16
-0.18
-0.20
-0.22
-0.24
-0.26
PCG12
0.02
-0.15
-0.17
9PCG
0.06
0.04
PCG
● Xyela sp
Xyela curva
Megaxyela euchroma
● Megalodontes spiraeae
Megalodontes quinquecinctus
Megalodontes cephalotes
Pamphilius sp
Chinolyda flagellicornis
● Pachycephus smyrnensis
Pachycephus cruentatus
Characopygus scythicus
Trachelus iudaicus
Trachelus tabidus
Cephus sareptanus
Cephus pygmeus
Cephus cinctus
Calameuta idolon
Calameuta filiformis
Hartigia linearis
Syrista parreyssii
● Perga condei
Arge similis
Arge bella
● Trichiosoma anthracinum
Labriocimbex sinicus
Corynis lateralis
● Allantus luctifer
Asiemphytus rufocephalus
Birmella discoidalisa
Analcellicampa xanthosoma
Monocellicampa pruni
Tenthredo tienmushana
Cladiucha punctata
Cladiucha magnoliae
● Orussus occidentalis
Xiphydria sp
Tremex columba
Taeniogonalos taihorina
Parapolybia crocea
● Mecoptera Neopanorpa phlchra
Diptera Anopheles gambiae
Megaloptera Neochauliodes parasparsus
Coleoptera Paroster microsturtensis

Fig. 3

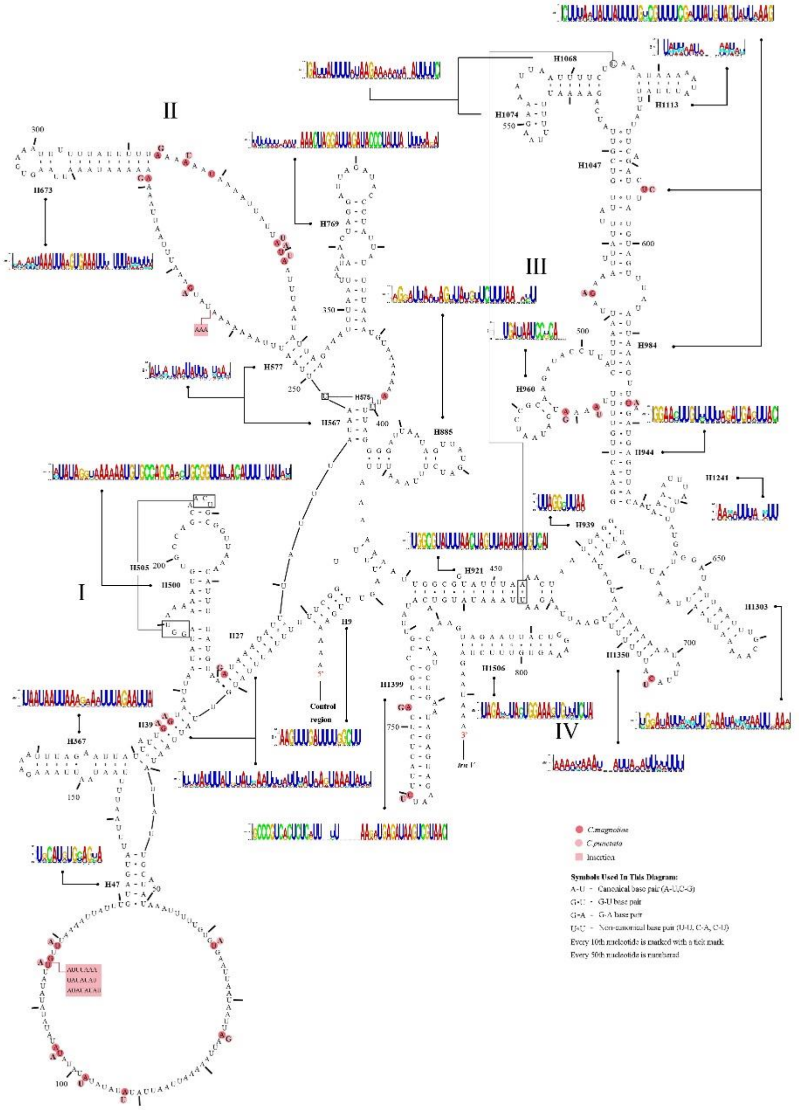

I
II
III
IV
H47
H139
H127
H9
H567
H500
H505
H577
H575
H673
H769
H885
H921
H939
H944
H960
H984
H1047
H1068
H1074
H1113
H1241
H1303
H1350
H1399
H1506
Control region
5'
3'
trnV
C.magnoliae
C.punctata
Insertion
Symbols Used In This Diagram:
A-U – Canonical base pair (A-U,C-G)
G•U – G-U base pair
G•A – G-A base pair
U•U – Non-canonical base pair (U-U, C-A, C-U)
Every 10th nucleotide is marked with a tick mark.
Every 50th nucleotide is numbered

Fig. 4

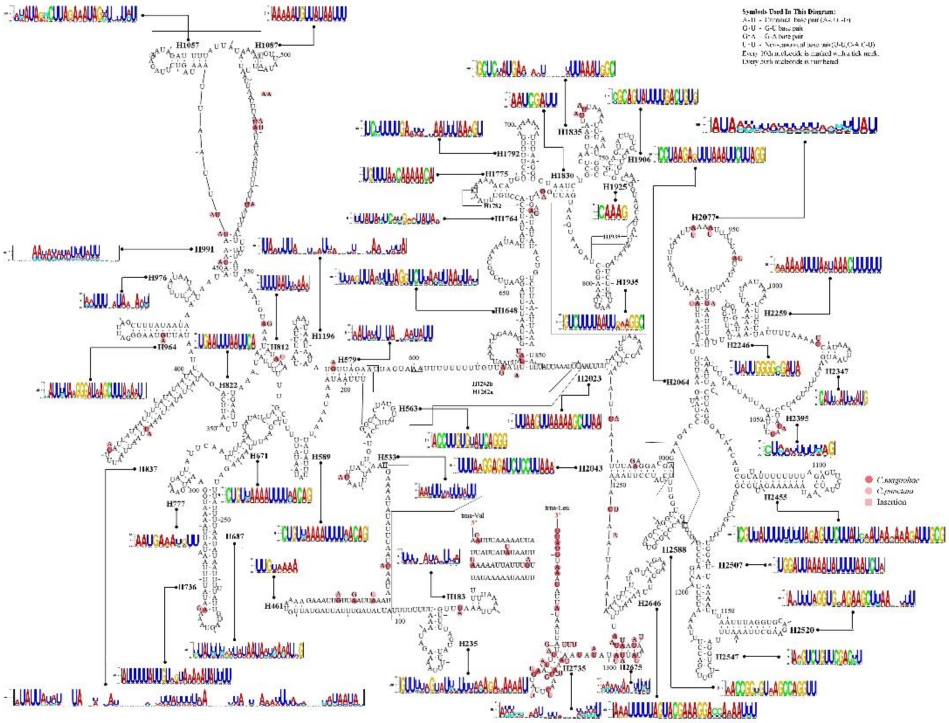

Symbols Used In This Diagram:
A-U – Canonical base pair (A-U, C-G)
G·U – G-U base pair
G·A – G-A base pair
U·U – Non-canonical base pair (U-U, C-A, C-U)
Every 10th nucleotide is marked with a tick mark.
Every 50th nucleotide is numbered
C.magnoliae
C.punctata
Insertion
H1057
H1087
H991
H976
H964
H812
H822
H837
H1196
H579
H563
H533
H671
H589
H777
H687
H736
H461
H183
H235
H1792
H1775
H1764
H1648
H1835
H1830
H1906
H1925
H1935
H2023
H2043
H2077
H2064
H2259
H2246
H2347
H2395
H2455
H2507
H2520
H2547
H2588
H2646
H2675
H2735
trna-Val
trna-Leu

Fig. 5

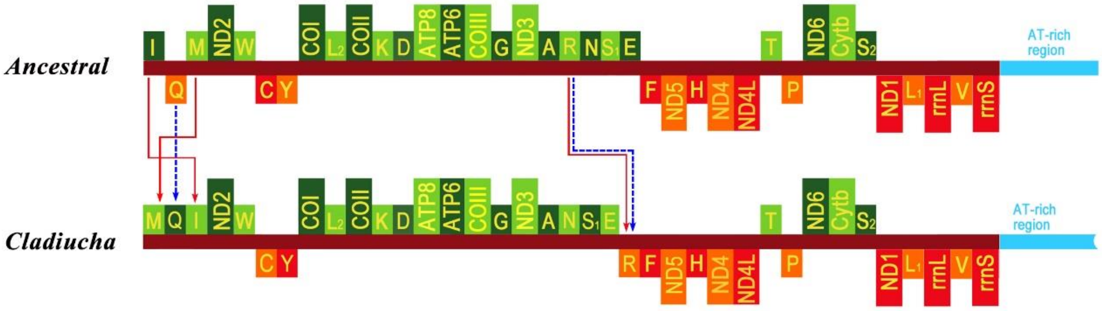

Ancestral
I
M
ND2
W
COI
L2
COII
K
D
ATP8
ATP6
COIII
G
ND3
A
R
N
S1
E
T
ND6
Cytb
S2
AT-rich region
Q
C
Y
F
ND5
H
ND4
ND4L
P
ND1
L1
rrnL
V
rrnS
Cladiucha
M
Q
I
ND2
W
COI
L2
COII
K
D
ATP8
ATP6
COIII
G
ND3
A
N
S1
E
T
ND6
Cytb
S2
AT-rich region
C
Y
R
F
ND5
H
ND4
ND4L
P
ND1
L1
rrnL
V
rrnS

Fig. 6

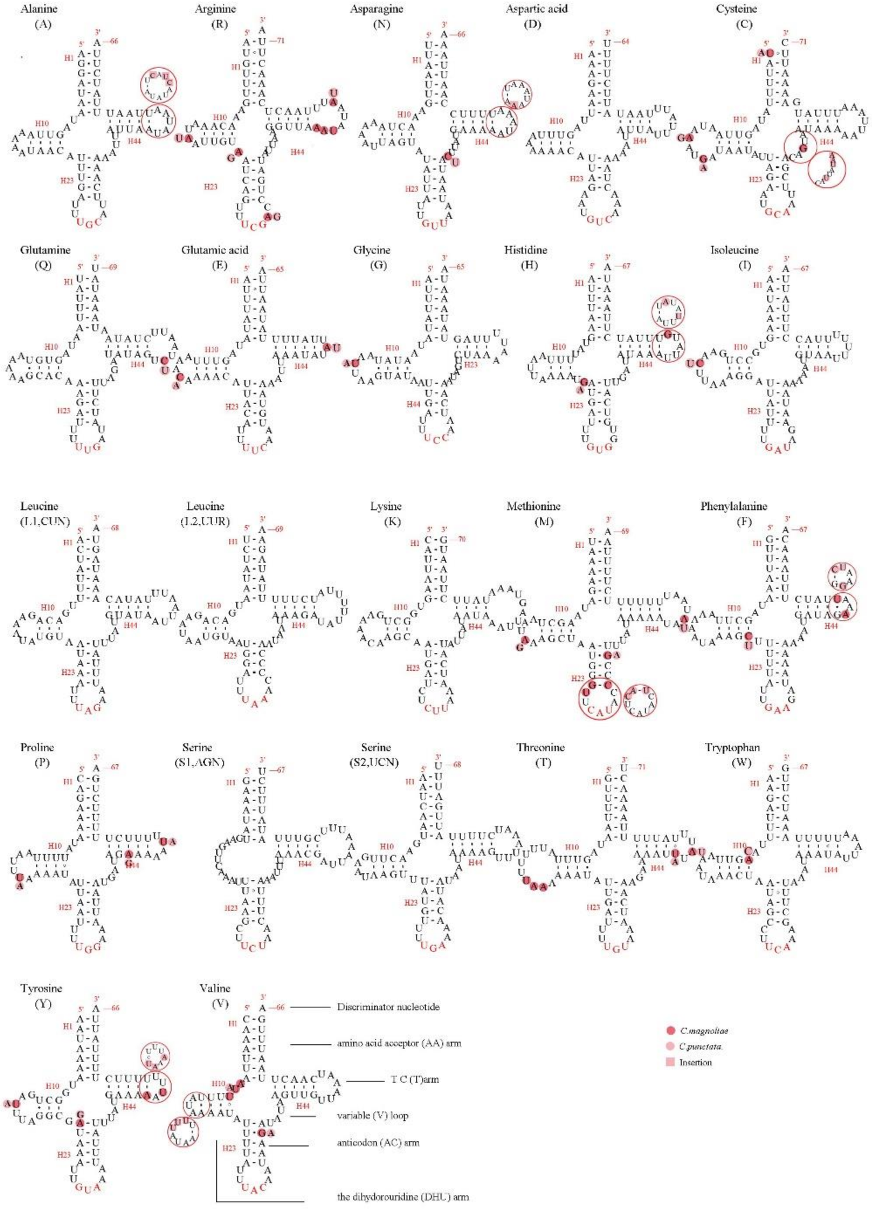

Alanine (A)
Arginine (R)
Asparagine (N)
Aspartic acid (D)
Cysteine (C)
Glutamine (Q)
Glutamic acid (E)
Glycine (G)
Histidine (H)
Isoleucine (I)
Leucine (L1,CUN)
Leucine (L2,UUR)
Lysine (K)
Methionine (M)
Phenylalanine (F)
Proline (P)
Serine (S1,AGN)
Serine (S2,UCN)
Threonine (T)
Tryptophan (W)
Tyrosine (Y)
Valine (V)
Discriminator nucleotide
amino acid acceptor (AA) arm
TΨC (T)arm
variable (V) loop
anticodon (AC) arm
the dihydorouridine (DHU) arm
C.magnoliae
C.punctata.
Insertion
H1
H10
H23
H44

Fig. 7

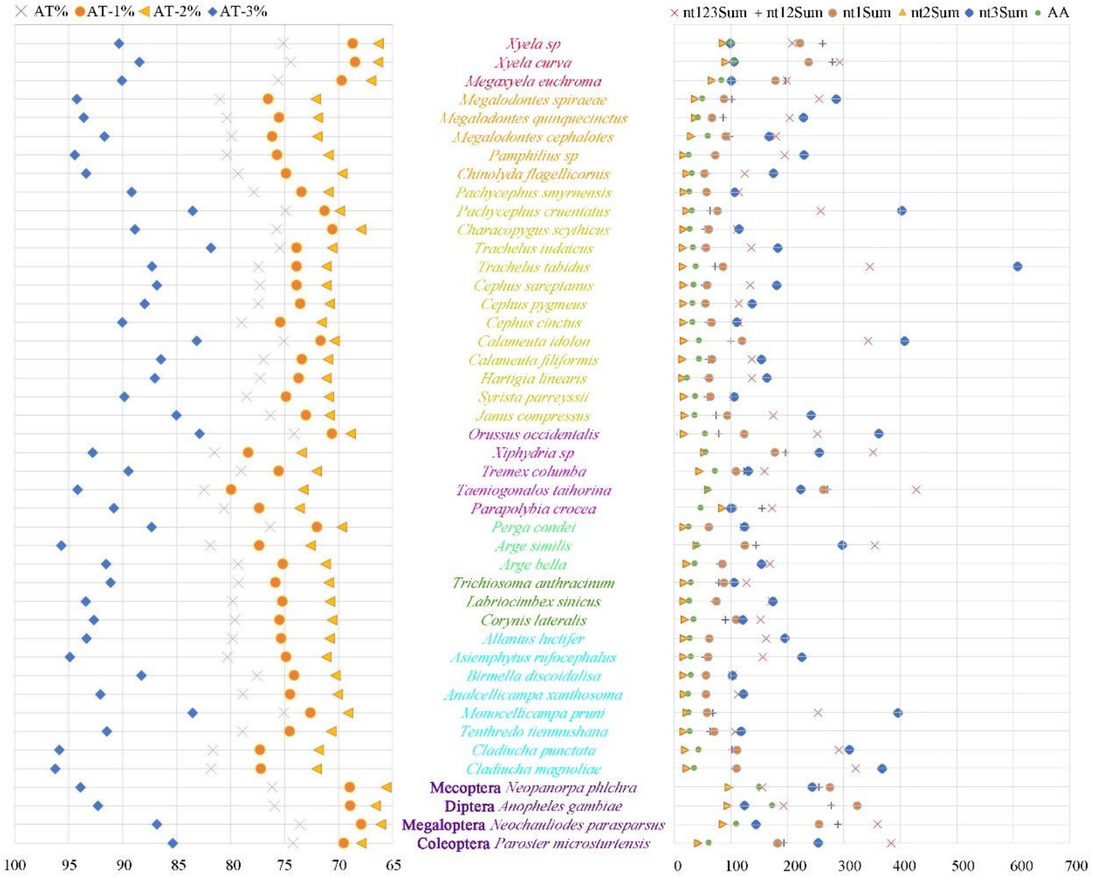

× AT% ● AT-1% ◀ AT-2% ◆ AT-3%
× nt123Sum + nt12Sum ● nt1Sum ▲ nt2Sum ● nt3Sum ● AA
Xyela sp
Xyela curva
Megaxyela euchroma
Megalodontes spiraeae
Megalodontes quinquecinctus
Megalodontes cephalotes
Pamphilius sp
Chinolyda flagellicornis
Pachycephus smyrnensis
Pachycephus cruentatus
Characopygus scythicus
Trachelus iudaicus
Trachelus tabidus
Cephus sareptanus
Cephus pygmeus
Cephus cinctus
Calameuta idolon
Calameuta filiformis
Hartigia linearis
Syrista parreyssii
Janus compressus
Orussus occidentalis
Xiphydria sp
Tremex columba
Taeniogonalos taihorina
Parapolybia crocea
Perga condei
Arge similis
Arge bella
Trichiosoma anthracinum
Labriocimbex sinicus
Corynis lateralis
Allantus luctifer
Asiemphytus rufocephalus
Birmella discoidalisa
Analcellicampa xanthosoma
Monocellicampa pruni
Tenthredo tienmushana
Cladiucha punctata
Cladiucha magnoliae
Mecoptera *Neopanorpa phlchra*
Diptera *Anopheles gambiae*
Megaloptera *Neochauliodes parasparsus*
Coleoptera *Paroster microsturtensis*
100
95
90
85
80
75
70
65
0
100
200
300
400
500
600
700

Fig. 8

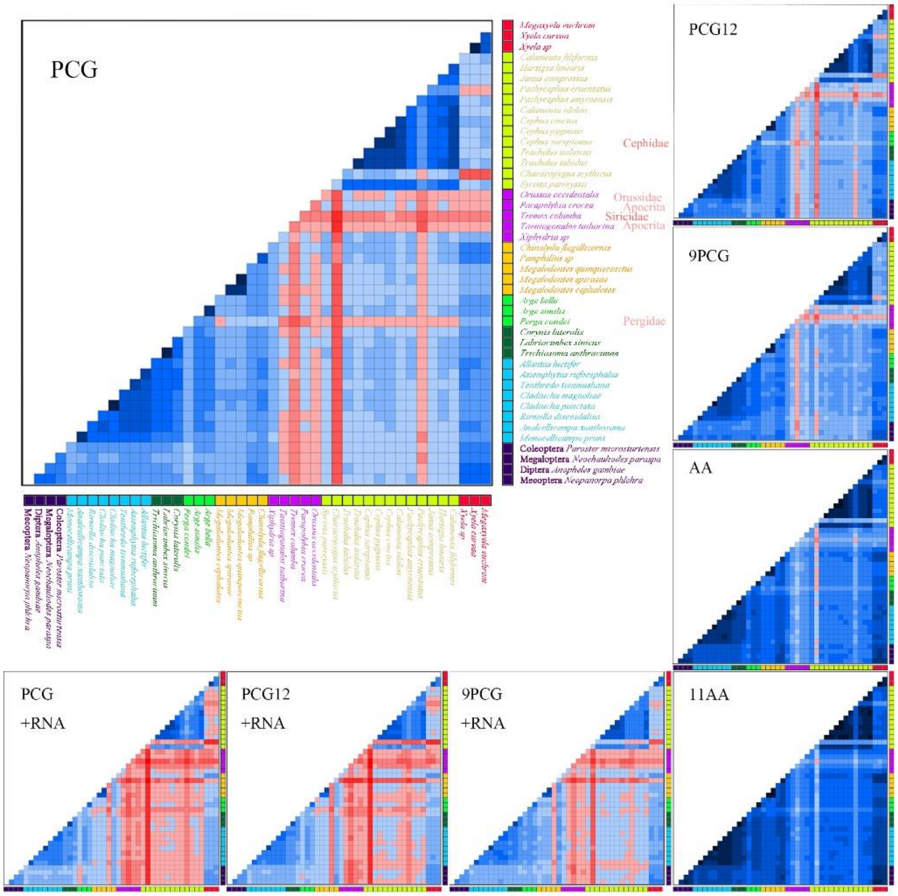

PCG
PCG12
9PCG
AA
PCG +RNA
PCG12 +RNA
9PCG +RNA
11AA
Megaxyela euchrom
Xyela curvaa
Xyela sp
Calameuta filiformis
Hartigia linearis
Janus compressus
Pachycephus cruentatus
Pachycephus smyrnensis
Calameuta idolon
Cephus cinctus
Cephus pygmeus
Cephus sareptanus
Trachelus iudaicus
Trachelus tabidus
Characopygus scythicus
Syrista parreyssii
Orussus occidentalis
Parapolybia crocea
Tremex columba
Taeniogonalos taihorina
Xiphydria sp
Chinolyda flagellicornis
Pamphilius sp
Megalodontes quinquecinctus
Megalodontes spiraeae
Megalodontes cephalotes
Arge bella
Arge similis
Perga condei
Corynis lateralis
Labriocimbex sinicus
Trichiosoma anthracinum
Allantus luctifer
Asiemphytus rufocephalus
Tenthredo tienmushana
Cladiucha magnoliae
Cladiucha punctata
Birmella discoidalisa
Analcellicampa xanthosoma
Monocellicampa pruni
Coleoptera Paroster microsturtensis
Megaloptera Neochauliodes parasparus
Diptera Anopheles gambiae
Mecoptera Neopanorpa phlchra
Cephidae
Orussidae
Apocrita
Siricidae
Apocrita
Pergidae

Fig. 9

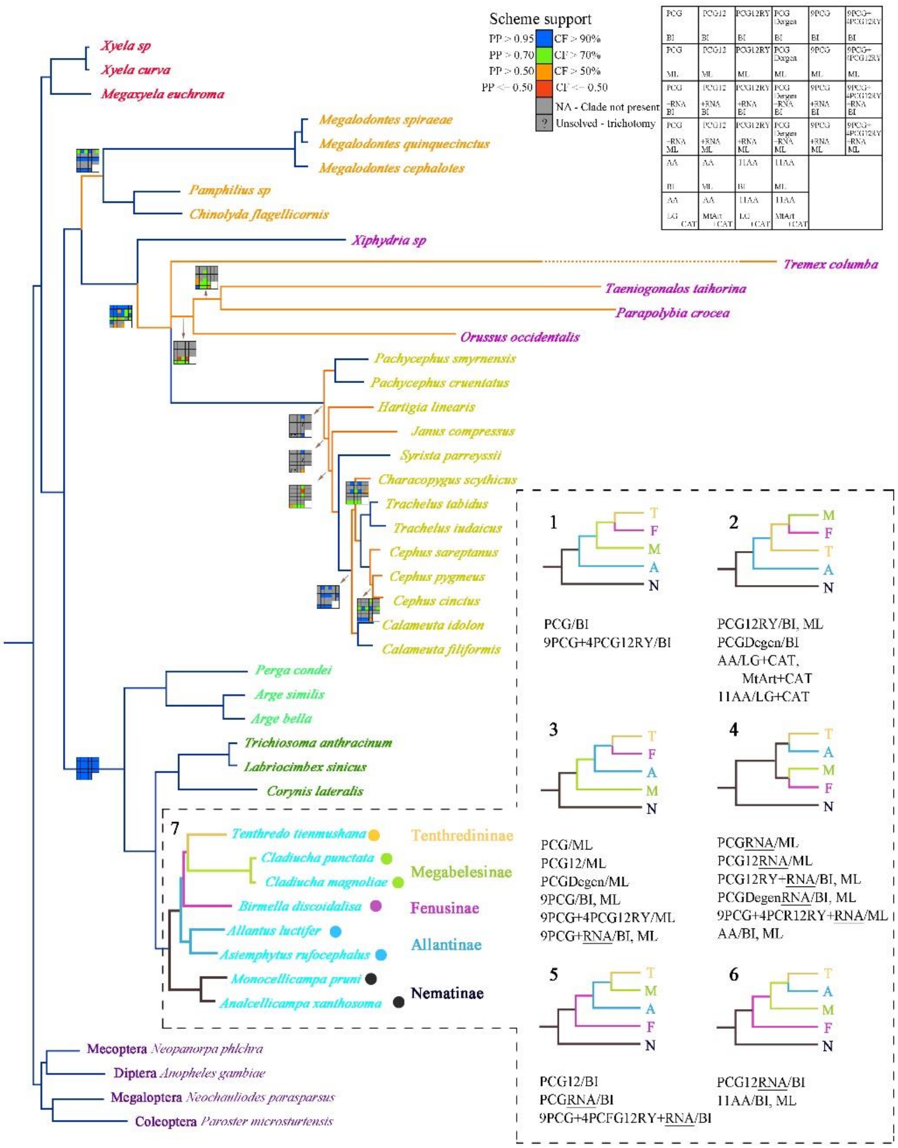

Scheme support
PP > 0.95 CF > 90%
PP > 0.70 CF > 70%
PP > 0.50 CF > 50%
PP <= 0.50 CF <= 0.50
NA - Clade not present
? Unsolved - trichotomy
PCG BI
PCG12 BI
PCG12RY BI
PCG Degen BI
9PCG BI
9PCG+4PCG12RY BI
PCG ML
PCG12 ML
PCG12RY ML
PCG Degen ML
9PCG ML
9PCG+4PCG12RY ML
PCG +RNA BI
PCG12 +RNA BI
PCG12RY +RNA BI
PCG Degen +RNA BI
9PCG +RNA BI
9PCG+4PCG12RY +RNA BI
PCG +RNA ML
PCG12 +RNA ML
PCG12RY +RNA ML
PCG Degen +RNA ML
9PCG +RNA ML
9PCG+4PCG12RY +RNA ML
AA BI
AA ML
11AA BI
11AA ML
AA LG +CAT
AA MtArt +CAT
11AA LG +CAT
11AA MtArt +CAT
Xyela sp
Xyela curva
Megaxyela euchroma
Megalodontes spiraeae
Megalodontes quinquecinctus
Megalodontes cephalotes
Pamphilius sp
Chinolyda flagellicornis
Xiphydria sp
Tremex columba
Taeniogonalos taihorina
Parapolybia crocea
Orussus occidentalis
Pachycephus smyrnensis
Pachycephus cruentatus
Hartigia linearis
Janus compressus
Syrista parreyssii
Characopygus scythicus
Trachelus tabidus
Trachelus iudaicus
Cephus sareptanus
Cephus pygmeus
Cephus cinctus
Calameuta idolon
Calameuta filiformis
Perga condei
Arge similis
Arge bella
Trichiosoma anthracinum
Labriocimbex sinicus
Corynis lateralis
7
Tenthredo tienmushana
Cladiucha punctata
Cladiucha magnoliae
Birmella discoidalisa
Allantus luctifer
Asiemphytus rufocephalus
Monocellicampa pruni
Analcellicampa xanthosoma
Tenthredininae
Megabelesinae
Fenusinae
Allantinae
Nematinae
Mecoptera Neopanorpa phlchra
Diptera Anopheles gambiae
Megaloptera Neochauliodes parasparsus
Coleoptera Paroster microsturtensis
1
T
F
M
A
N
PCG/BI
9PCG+4PCG12RY/BI
2
M
F
T
A
N
PCG12RY/BI, ML
PCGDegen/BI
AA/LG+CAT,
MtArt+CAT
11AA/LG+CAT
3
T
F
A
M
N
PCG/ML
PCG12/ML
PCGDegen/ML
9PCG/BI, ML
9PCG+4PCG12RY/ML
9PCG+RNA/BI, ML
4
T
A
M
F
N
PCGRNA/ML
PCG12RNA/ML
PCG12RY+RNA/BI, ML
PCGDegenRNA/BI, ML
9PCG+4PCR12RY+RNA/ML
AA/BI, ML
5
T
M
A
F
N
PCG12/BI
PCGRNA/BI
9PCG+4PCFG12RY+RNA/BI
6
T
A
M
F
N
PCG12RNA/BI
11AA/BI, ML

Fig. 10

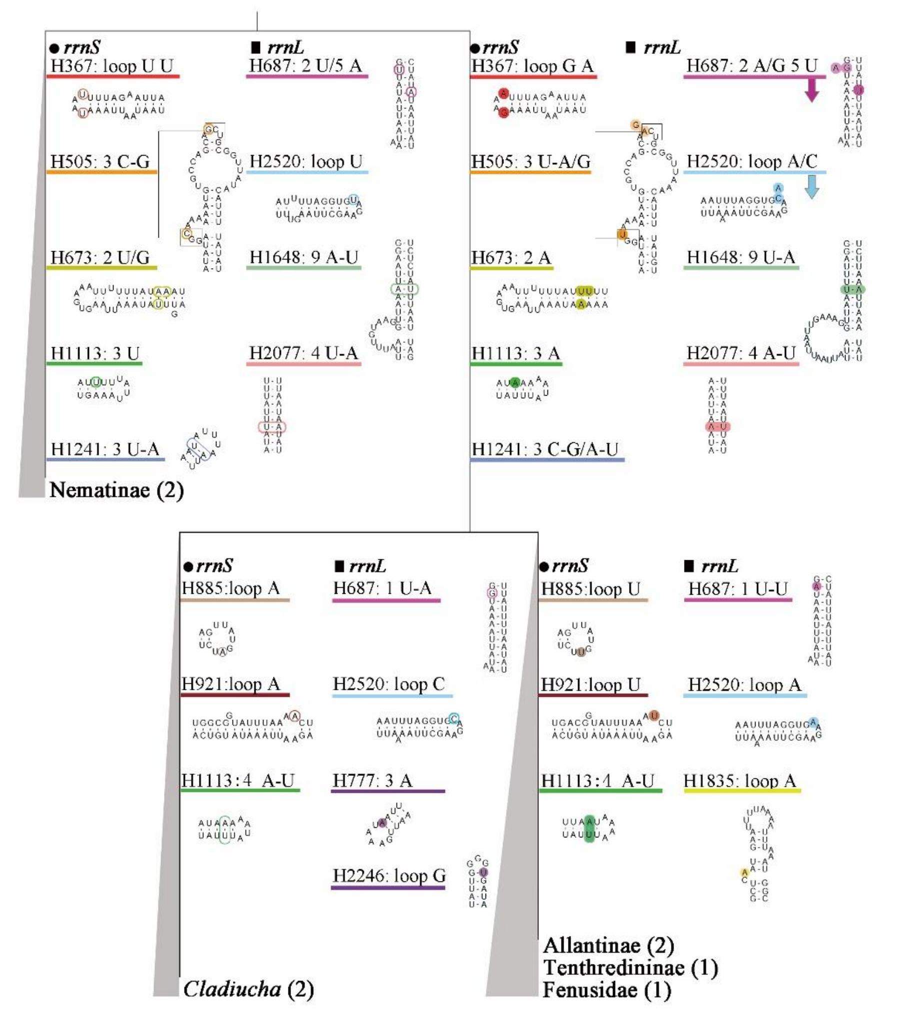
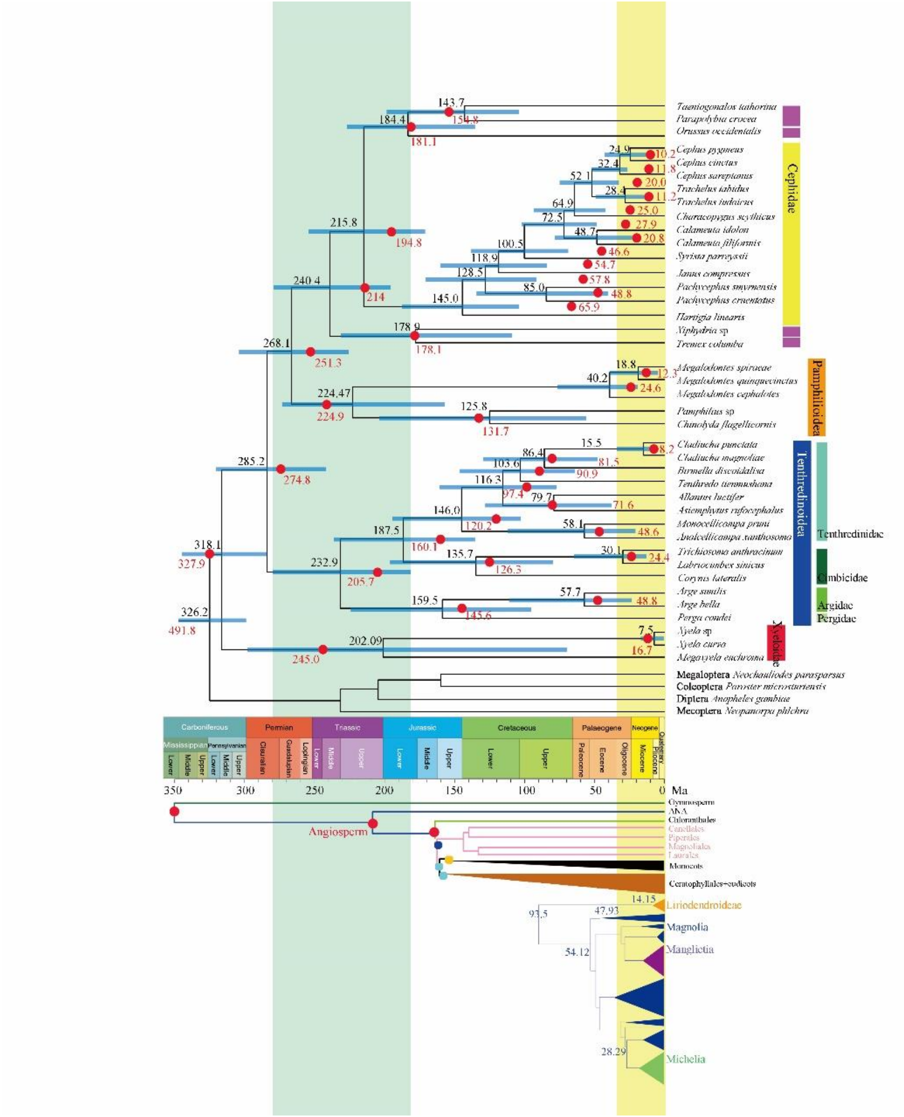

● rrnS
H367: loop U U
H505: 3 C-G
H673: 2 U/G
H1113: 3 U
H1241: 3 U-A
■ rrnL
H687: 2 U/5 A
H2520: loop U
H1648: 9 A-U
H2077: 4 U-A
Nematinae (2)
● rrnS
H367: loop G A
H505: 3 U-A/G
H673: 2 A
H1113: 3 A
H1241: 3 C-G/A-U
■ rrnL
H687: 2 A/G 5 U
H2520: loop A/C
H1648: 9 U-A
H2077: 4 A-U
● rrnS
H885:loop A
H921:loop A
H1113: 4 A-U
■ rrnL
H687: 1 U-A
H2520: loop C
H777: 3 A
H2246: loop G
Cladiucha (2)
● rrnS
H885:loop U
H921:loop U
H1113: 4 A-U
■ rrnL
H687: 1 U-U
H2520: loop A
H1835: loop A
Allantinae (2)
Tenthredininae (1)
Fenusidae (1)

Fig. 11



Taeniogonalos taihorina
Parapolybia crocea
Orussus occidentalis
Cephus pygmeus
Cephus cinctus
Cephus sareptanus
Trachelus tabidus
Trachelus iudaicus
Characopygus scythicus
Calameuta idolon
Calameuta filiformis
Syrista parreyssii
Janus compressus
Pachycephus smyrnensis
Pachycephus cruentatus
Hartigia linearis
Xiphydria sp
Tremex columba
Cephidae
Megalodontes spiraeae
Megalodontes quinquecinctus
Megalodontes cephalotes
Pamphilius sp
Chinolyda flagellicornis
Pamphilioidea
Cladiucha punctata
Cladiucha magnoliae
Birmella discoidalisa
Tenthredo tienmushana
Allantus luctifer
Asiemphytus rufocephalus
Monocellicampa pruni
Analcellicampa xanthosoma
Trichiosoma anthracinum
Labriocimbex sinicus
Corynis lateralis
Arge similis
Arge bella
Perga condei
Tenthredinoidea
Tenthredinidae
Cimbicidae
Argidae
Pergidae
Xyela sp
Xyela curva
Megaxyela euchroma
Xyelidae
Megaloptera Neochauliodes parasparsus
Coleoptera Pārvister microsturtensis
Diptera Anopheles gambiae
Mecoptera Neopanorpa phlchra
143.7
154.8
184.4
181.1
24.9
10.2
32.4
11.8
52.1
20.0
28.4
11.2
64.9
25.0
72.5
27.9
48.7
20.8
100.5
46.6
118.9
54.7
128.5
57.8
85.0
48.8
145.0
65.9
215.8
194.8
240.4
214
178.9
178.1
268.1
251.3
18.8
12.3
40.2
24.6
224.47
224.9
125.8
131.7
15.5
8.2
86.4
81.5
103.6
90.9
116.3
97.4
79.7
71.6
146.0
120.2
58.1
48.6
187.5
160.1
135.7
126.3
30.1
24.4
232.9
205.7
57.7
48.8
159.5
145.6
285.2
274.8
318.1
327.9
326.2
491.8
202.09
245.0
7.5
16.7
Carboniferous
Permian
Triassic
Jurassic
Cretaceous
Palaeogene
Neogene
Quaternary
Mississippian
Pennsylvanian
Lower
Middle
Upper
Lower
Middle
Upper
Cisuralian
Guadalupian
Lopingian
Lower
Middle
Upper
Lower
Middle
Upper
Lower
Upper
Paleocene
Eocene
Oligocene
Miocene
Pliocene
350
300
250
200
150
100
50
0
Ma
Gymnosperm
ANA
Chloranthales
Canellales
Piperales
Magnoliales
Laurales
Monocots
Ceratophyllales+eudicots
Angiosperm
14.15
Liriodendroideae
93.5
47.93
Magnolia
54.12
Manglietia
28.29
Michelia

Tables

Supplementary Table 1 Summary information of symphytan mitochondrial genomes used in phylogenetic analyses

	Species	**Family**	**Accesion number**	**References**
Ingroup	*Cladiucha magnoliae*	Tenthredinidae		This study
	Cladiucha punctata	Tenthredinidae		This study
	Labriocimbex sinica	Cimbicidae	MH136623	Yan et al, 2019
	Corynis lateralis	Cimbicidae	KY063728	Doğan and Korkmaz, 2017
	Trichiosoma anthracinum	Cimbicidae	KT921411	Song *et al*., 2016
	Megalodontes cephalotes	Megalodontesidae	MH577058	Niu *et al*., 2018
	Megalodontes spiraeae	Megalodontesidae	MH577059	Niu *et al*., 2018
	Megalodontes quinquecinctus	Megalodontesidae	MG923502	Tang *et al*., 2019
	Analcellicampa xanthosoma	Tenthredinidae	MH992752	Unpublished
	Allantus luctifer	Tenthredinidae	KJ713152	Wei *et al*., 2014
	Asiemphytus rufocephalus	Tenthredinidae	KR703582	Song *et al*., 2016
	Monocellicampa pruni	Tenthredinidae	JX566509	Wei *et al*., 2013
	Tenthredo tienmushana	Tenthredinidae	KR703581	Song *et al*., 2015
	Birmella discoidalisa	Tenthredinidae	MF197548	Unpublished
	Xyela sp.	Xyelidae	MG923517	Tang *et al*., 2019
	Xyela curva	Xyelidae	unpulished	unpulished
	Megaxyela euchroma	Xyelidae	unpulished	unpulished
	Xiphydria sp.	Xiphydriidae	MH422969	Ma *et al*., 2018
	Tremex columba	Siricidae	MH422968	Ma *et al*., 2018
	Pamphilius sp.	Pamphiliidae	MG923504	Tang *et al*., 2019
	Chinolyda flagellicornis	Pamphiliidae	MH577057	Niu *et al*., 2018
	Orussus occidentalis	Orussidae	FJ478174	Dowton *et al*., 2009
	Arge similis	Argidae	MG923484	Tang *et al*., 2019
	Arge bella	Argidae	MF287761	Du *et al*., 2018
	Calameuta filiformis	Cephidae	KT260167	Korkmaz *et al*., 2016
	Calameuta idolon	Cephidae	KT260168	Korkmaz *et al*., 2016
	Cephus cinctus	Cephidae	FJ478173	Dowton *et al*., 2009
	Cephus pygmeus	Cephidae	KM377623	Korkmaz *et al*., 2015
	Cephus sareptanus	Cephidae	KM377624	Korkmaz *et al*., 2015
	Characopygus scythicus	Cephidae	KX907848	Korkmaz *et al*., 2018
	Hartigia linearis	Cephidae	KX907843	Korkmaz *et al*., 2018
	Janus compressus	Cephidae	KX907844	Korkmaz *et al*., 2018
	Pachycephus cruentatus	Cephidae	KX907845	Korkmaz *et al*., 2018
	Pachycephus smyrnensis	Cephidae	KX907846	Korkmaz *et al*., 2018
	Syrista parreyssi	Cephidae	KX907847	Korkmaz *et al*., 2018
	Trachelus iudaicus	Cephidae	KX257357	Korkmaz *et al*., 2017
	Trachelus tabidus	Cephidae	KX257358	Korkmaz *et al*., 2017
	Perga condei	Pergidae	AY787816	Castro and Dowton, 2005
	Taeniogonalos taihorina	Trigonalidae	NC027830	Wu *et al*., 2014
	Parapolybia crocea	Vespidae	KY679828	Peng *et al*., 2017
Outgroup	*Paroster microsturtensis*	Dytiscidae	MG912997	Hyde *et al*., 2018
	Neopanorpa phlchra	Panorpidae	FJ169955	Unpublished
	Neochauliodes parasparsus	Corydalidae	KX821680	Zhao *et al*., 2017
	Anopheles gambiae	Culicidae	L20934	Beard *et al*., 1993

Table 1 Summary of mitochondrial genome of *Cladiucha* magnoliae and C. punctata

ne	Strand	*Cladiucha* magnoliae							*Cladiucha* punctata						
		Start	Stop	Length(bp)	Start codon	Stop codon	Anticodon	IGN	Start	Stop	Length(bp)	Start codon	Stop codon	Anticodon	IGN
trnM	J	1	69	69			CAT		1	69	69			CAT	
trnQ	J	79	147	69			TTG	9	79	147	69			TTG	9
trnI	J	156	222	67			GAT	8	170	236	67			GAT	22
nad2	J	248	1,300	1,053	ATG	TAA		25	287	1,336	1,050	ATG	TAA		50
trnW	J	1,306	1,372	67			TCA	5	1,351	1,417	67			TCA	14
trnC	N	1,372	1,443	72			GCA	-1	1,417	1,491	75			GCA	-1
trnY	N	1,462	1,526	65			GTA	18	1,501	1,566	66			GTA	9
cox1	J	1,535	3,073	1,539	ATT	TAA		8	1,573	3,108	1,536	ATT	TAA		6
trnL2	J	3,100	3,168	69			TAA	26	3,154	3,222	69			TAA	45
cox2	J	3,169	3,849	681	ATT	TAA		0	3,223	3,903	681	ATT	TAA		0
trnK	J	3,854	3,923	70			CTT	4	3,906	3,975	70			CTT	2
trnD	J	3,924	3,987	64			GTC	0	3,976	4,039	64			GTC	0
atp8	J	3,988	4,149	162	ATT	TAA		0	4,040	4,201	162	ATT	TAA		0
atp6	J	4,143	4,820	678	ATG	TAA		-7	4,195	4,872	678	ATG	TAA		-7
cox3	J	4,820	5,608	789	ATG	TAA		-1	4,872	5,657	786	ATG	TAA		-1
trnG	J	5,614	5,678	65			TCC	5	5,663	5,727	65			TCC	5
nad3	J	5,679	6,032	354	ATA	TAA		0	5,728	6,081	354	ATA	TAA		0
trnA	J	6,041	6,106	66			TGC	8	6,099	6,167	69			TGC	17
trnN	J	6,146	6,213	68			GTT	39	6,207	6,275	69			GTT	39
trnS1	J	6,214	6,280	67			GCT	0	6,276	6,342	67			GCT	0
trnE	J	6,288	6,352	65			TTC	7	6,350	6,414	65			TTC	7
trnR	N	6,351	6,421	71			TCG	-2	6,413	6,480	68			TCG	-2
trnF	N	6,429	6,495	67			GAA	7	6,488	6,555	68			GAA	7
nad5	N	6,526	8,244	1,719	ATT	TAA		30	6,578	8,296	1,719	ATT	TAA		22
trnH	N	8,245	8,311	67			GTG	0	8,297	8,364	68			GTG	0
nad4	N	8,312	9,656	1,345	ATA	T		0	8,365	9,709	1,345	ATA	T		0
nad4L	N	9,653	9,946	294	ATT	TAA		-4	9,706	9,999	294	ATT	TAA		-4
trnT	J	9,949	10,019	71			TGT	2	10,002	10,069	68			TGT	2
trnP	N	10,020	10,086	67			TGG	0	10,070	10,136	67			TGG	0
ND6	J	10,088	10,600	513	ATA	TAA		1	10,138	10,650	513	ATA	TAA		1
cob	J	10,600	11,733	1,134	ATG	TAA		-1	10,650	11,783	1,134	ATG	TAA		-1
trnS2	J	11,734	11,801	68			TGA	0	11,784	11,851	68			TGA	0
nad1	N	11,840	12,793	954	ATA	TAA		38	11,886	12,839	954	ATA	TAA		34
trnL1	N	12,794	12,861	68			TAG	0	12,840	12,907	68			TAG	0
rrnL	N	12,862	14,205	1,344				0	12,908	14,248	1,341				0
trnV	N	14,206	14,271	66			TAC	0	14,249	14,315	67			TAC	0
rrnS	N	14,402	15,214	813				130	14,467	15,301	835				151